

Article

Linear Response of General Observables in Spiking Neuronal Network Models

Bruno Cessac ¹, Ignacio Ampuero ² and Rodrigo Cofré ^{3,*} 

¹ Biovision Team, INRIA and Neuromod Institute, Université Côte d'Azur, 06902 Sophia Antipolis, France; bruno.cessac@inria.fr

² Departamento de Informática, Universidad Técnica Federico Santa María, Valparaíso 2340000, Chile; ignacio.ampuero.13@sansano.usm.cl

³ CIMFAV-Ingemat, Facultad de Ingeniería Universidad de Valparaíso, Valparaíso 2340000, Chile

* Correspondence: rodrigo.cofre@uv.cl

Abstract: We establish a general linear response relation for spiking neuronal networks, based on chains with unbounded memory. This relation allow us to predict the influence of a weak amplitude time dependent external stimuli on spatio-temporal spike correlations, from the spontaneous statistics (without stimulus) in a general context where the memory in spike dynamics can extend arbitrarily far in the past. Using this approach, we show how the linear response is explicitly related to the collective effect of the stimuli, intrinsic neuronal dynamics, and network connectivity on spike train statistics. We illustrate our results with numerical simulations performed over a discrete time integrate and fire model.

Keywords: neuronal network dynamics; spike train statistics; linear response; non-Markovian dynamics; Gibbs distributions; maximum entropy principle



Citation: Cessac, B.; Ampuero, I.; Cofré, R. Linear Response of General Observables in Spiking Neuronal Network Models. *Entropy* **2021**, *23*, 155. <https://doi.org/10.3390/e23020155>

Academic Editor: David Holcman
Received: 10 December 2020
Accepted: 21 January 2021
Published: 27 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Neurons communicate by short-lasting electrical signals called action potentials or “spikes”, allowing the rapid propagation of information throughout the nervous system, with a minimal energy dissipation [1]. The spike shape is remarkably constant for a given neuron, and it is a contemporary view to consider spikes as quanta (bits) of information [2]. As a result, information is presumably encoded in the spike timing [3].

The simplest quantitative way to characterize the spiking activity of a neuron is its firing rate $r(t)$, where $r(t)dt$ is the probability that this neuron spikes during a small interval $[t, t + dt]$. Under the influence of an external stimulus the firing rate changes. A classical ansatz, coming from the Volterra expansions [2] is to write the variation in the firing rate of a neuron as the convolution form:

$$\delta^{(1)}[r(t)] = (K * S)[t], \quad (1)$$

where the exponent ⁽¹⁾ recalls that we consider a first-order effect of the stimulus S , that is, the stimulus is weak enough so that higher-order terms in the Volterra expansion can be neglected. This is an example of linear response: the variation in the rate is proportional to the stimulus. Here, K is a convolution kernel constrained by the underlying network dynamics. For example, in sensory neurons, where S and K are functions of space and time, the convolution (1) takes the explicit form:

$$(K * S)[t] = \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \int_{\tau=-\infty}^t K(x, y, t - \tau) S(x, y, \tau) d\tau dx dy, \quad (2)$$

where K decays sufficiently fast at infinity (in space and time) to ensure that the integral is well defined. K mimics the receptive field (RF) of the neuron. In general, the response

of spiking neuronal networks to a time-dependent stimulus does not only affect rates, it also has an impact on higher-order correlations between neurons, because neurons are connected through synapses. This situation is sketched in Figure 1.

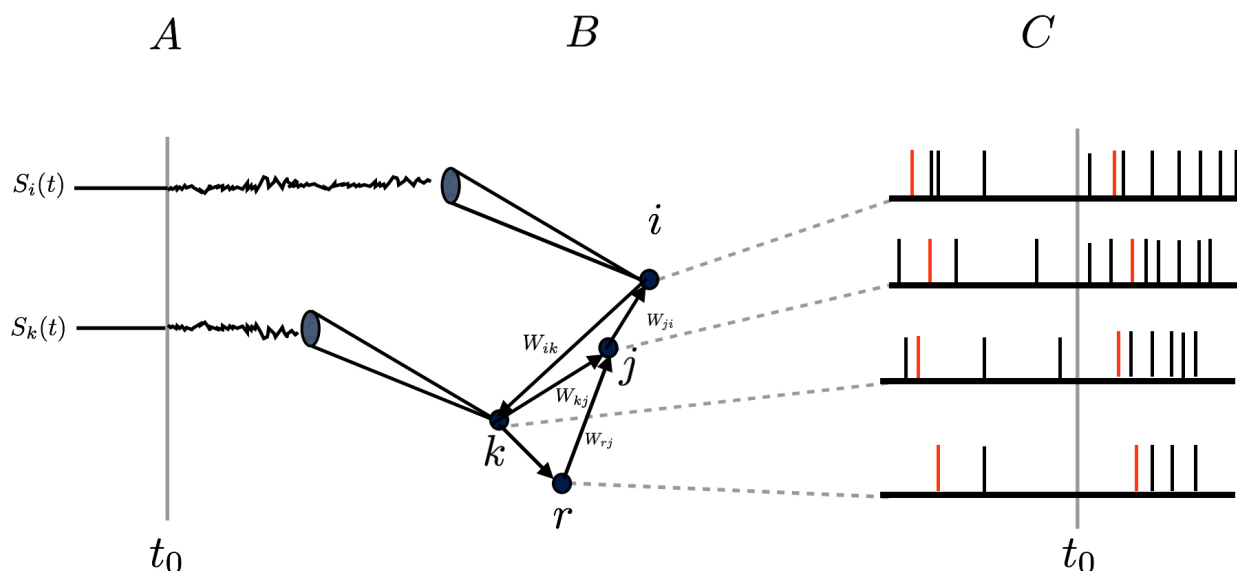


Figure 1. A time-dependent stimulus (A) is applied at time t_0 to some neurons of a spiking neuronal network (B). As a result, the spiking activity is modified as well as spike correlations between neurons (C), even for neurons not directly stimulated, because of direct or indirect synaptic interactions. These are represented in the figure by the weights W .

In particular, sensory neurons collectively convey to the brain information about external stimuli using correlated spike patterns resulting from the conjunction of stimulus influence, intrinsic neurons dynamics and neurons interactions via synapses [4–8]. This correlated firing has been linked to stimulus encoding [9], stimulus discrimination [10,11] and to intrinsic properties of the network which remain in absence of stimulus [12]. However, disentangling the biophysical origins of the correlations observed in spiking data is still a central and difficult problem in neuroscience [2,13,14]. As a consequence, correlations in spiking neuronal networks have attracted a considerable amount of attention in the last years, from experimental data analysis perspectives [4,6,7,15] as well as from the theoretical modeling viewpoint [1,16–20].

On one hand, novel experimental recording techniques such as Multi-Electrode Arrays (MEA) permit the measurement of the collective spiking activity of larger and larger populations of interacting neurons responding to external stimuli [21,22]. These recordings allow, in particular, for better characterizations of the link between stimuli and the correlated responses of a living neuronal network, paving the way to better understand “the neural-code” [2]. Yet, neuronal responses are highly variable [13]. Even at the single neuron level, when presenting repetitions of the same stimulus under controlled experimental conditions, the neural activity changes from trial to trial [23,24]. Thus, researchers are seeking statistical regularities in order to unveil a probabilistic, causal, relationship between stimuli and spiking responses [4,5,7,25–27].

On the other hand, mathematical models of spiking neuronal networks offer a complementary approach to biological experiments [13,14,28,29]. Based on bio-physically plausible mechanisms describing the dynamics of neurons, mathematical modeling provides a framework to characterize the population spike train statistics in terms of biophysical parameters, synaptic connectivity, history of previous spikes and stimuli. The hope is that understanding these aspects in a model may help to better the processing and extraction of information from real data.

In this article, we consider a spiking neuronal network model where the non-linear dynamics and neurons interactions naturally produce spatio-temporal spikes correlations. We assume that these neurons reach a stationary state without stimuli and from a given time are submitted to a time-dependent stimulation (See Figure 1). How are the spatio-temporal spike correlations modified by this stimulus? We address this question in the context of linear response theory using methods from ergodic theory and so-called chains with complete connections [30], extending the notion of Markov chains to infinite memory, providing a generalized notion of Gibbs distribution. We show that spatio-temporal response is written in terms of a history-dependent convolution kernel applied to the stimuli, generalizing (2) to more general observables than rates. We compute explicitly this kernel in a specific model example and analyze the validity of our linear response formula by numerical means.

The linear response determines how the expectation value of an observable of a dynamical system changes upon weakly perturbing the dynamics. The response of a system, originally at equilibrium, to a time-dependent stimulus is proportional to the stimulus, with coefficients obtained via correlation functions computed at equilibrium. The theory we develop has its roots in non-equilibrium statistical physics (briefly reviewed in Section 2.1). A seminal result in this context is the fluctuation–dissipation theorem [31], where the linear response only depends on the correlation functions of the unperturbed system. Our approach proceeds along similar lines, meaning that the linear response can be predicted from the spikes correlations of the unperturbed system (stationary state). We derive a result of this type for a discrete-time integrate and fire model addressing a central question: which correlations matter and how they are related to synaptic interactions? Among the applications of our results is the characterization of population receptive fields in terms of the underlying synaptic connectivity, non-linear dynamics and background activity, the prediction of higher-order correlations in the population responses to non-stationary weak amplitude stimuli, assessing the role of the synaptic connectivity in motion processing from the spiking response to, etc.

The paper is organized as follows. In Section 2, we briefly review linear response in statistical physics and ergodic theory allowing us to make a link between neuronal networks, considered as dynamical systems, and the statistics of spikes. We introduce the formalism of chains with unbounded memory (which are, as we explain, equivalent to left-sided one-dimensional Gibbs distributions), allowing the handling of non-stationary spike distribution with unbounded memory. In Section 3, we derive the general linear response formula (16) used throughout the paper. This equation expresses the time-dependent variation in the average of an observable f as a time series of specific correlation functions computed with respect to spontaneous activity (without stimulus). This result, reminiscent of the fluctuation-dissipation theorem in statistical physics [32,33], is applied here to spike statistics. In Section 4 we introduce a spiking neuronal network model to instantiate our analysis. This model has been presented in [34]. We associate the spiking activity to a discrete stochastic process defined from transition probabilities where memory is unbounded. These probabilities are written as a function of the parameters of the model. We explicitly wrote a discrete-time form of the convolution kernel (1) as an explicit function of the model parameters, especially synaptic weights. The expression relies on a Markovian approximation of the chain and on a decomposition theorem of spike observables, introduced in a more general context by Hammersley and Clifford in 1971 [35] (see Section 2.2). In Section 5 we illustrate through an example our linear response theory applied. In particular, we show how one can predict the variation in the firing rate and in the delayed pairwise correlation between two neurons from the mere knowledge of the stimulus and relevant spontaneous correlations. Finally, in Section 6 we discuss our results.

2. Linear Response, Gibbs Distributions and Probabilistic Chains with Unbounded Memory

Neuronal networks can be considered either as dynamical systems (when the dynamics is known) or as spike generating processes characterized by transition probabilities computed from spike train observations. In the first case, it is natural to seek a linear response from dynamics itself, using approximations (e.g., mean-field [36]). In the second case, one has to define a probability distribution on the spike trains in order to investigate the effect of a perturbation. In this section, we show how these two approaches are related, making a link between the classical statistical physics approach of linear response, dynamical systems and ergodic theory, and neuronal networks. We introduce then the general formalism of chains with unbounded memory allowing the handling of non-equilibrium linear response for spiking neuronal networks. All of the material in this section is known in different domains, statistical physics, ergodic theory, stochastic processes, neuronal networks, and is presented here for a better understanding of the next sections.

2.1. Linear Response in Statistical Physics

For simplicity, we consider in this introductory section a dynamical system taking a finite number of “configurations” denoted by ω . In statistical physics, the linear response theory can be addressed in these terms. In a system at thermodynamic equilibrium, the probability of a configuration ω is given by the Boltzmann–Gibbs distribution:

$$\mu[\omega] = \frac{1}{Z} e^{-\frac{H(\omega)}{k_B T}}, \quad (3)$$

where $Z = \sum_{\omega} e^{-\frac{H(\omega)}{k_B T}}$ is called the partition function, with k_B , the Boltzmann constant and T the temperature. The function $H(\omega)$ is called the energy

$$H(\omega) = \sum_{\alpha} \lambda_{\alpha} X_{\alpha}(\omega). \quad (4)$$

The functions X_{α} are extensive quantities (proportional to the number of particles) such as energy, electric charge, volume, number of particles, magnetic field, ... The conjugated parameters λ_{α} correspond to intensive quantities (not proportional to the number of particles), like temperature, electric potential, pressure, chemical potential, magnetic susceptibility, ... In general, they depend on the location in the physical space (e.g., the temperature depends on the position in a fluid). At equilibrium, they are uniform in space though. The form of H , i.e., the choice of the λ_{α} and X_{α} is constrained by the physical properties of the system. It is also constrained by boundary conditions.

In standard statistical physics courses, the Gibbs distribution form (3) is obtained as a consequence of the Maximum Entropy Principle [37]. For a probability measure P on the set of configurations, the statistical entropy is:

$$S[P] = -k_B \sum_{\omega} \log P[\omega] \log P[\omega]. \quad (5)$$

Denote $E_P[\]$ the expectation with respect to P . The Maximum Entropy Principle seeks a probability distribution maximizing the statistical entropy under the constraint that the average energy is constant, i.e., $E_P[H] = C$ for any probability measure P on the set of configurations. This probability exists and is unique when the set of configurations is finite; this is (3). When this set is infinite (e.g., thermodynamic limit) additional summability conditions are required on H to ensure existence and uniqueness of P [38,39].

A non-equilibrium situation arises when the λ_{α} s are not uniform in space, generating gradients $\vec{\nabla} \lambda_{\alpha}$ (temperature gradient, electric potential gradient ...). These gradients result in currents of X_{α} (e.g., a temperature gradient induces a heat current). In general, the currents are nonlinear functions of gradients. The Onsager linear response theory assumes that currents are linear combinations of gradients (i.e., gradients are weak enough so that

non-linear terms can be neglected). Known examples are Ohm's law where the electric current is proportional to the gradient of the electric potential, Fourier's law where the heat flux is proportional to the temperature gradient, Fick's law etc. Several gradients can be simultaneously involved like in the Peltier effect. The proportionality coefficients are called Onsager coefficients [40].

Now, the property that interests us is that Onsager coefficients are obtained as correlations functions computed at equilibrium (Kubo relations [32]). Thus, the knowledge of correlations at equilibrium allows the inference of the non-equilibrium response of the system to weak perturbations.

2.2. Linear Response in Spiking Neuronal Networks

Neuronal networks are modeled by dynamical systems (possibly stochastic). Therefore, the linear response theory can be addressed using the tools briefly presented in the previous section. This has been done for discrete-time in the Amari–Wilson–Cowan model [41,42], where the convolution kernel K appearing in (2) can be explicitly computed [43,44]. Notably, one can compute the response of a neuron to a weak harmonic perturbation of another neuron, exhibiting specific resonances and functional connectivity distinct from the synaptic graph.

When dealing with spiking models such as the Integrate and Fire, the dynamics are not differentiable anymore (because of the mechanism of reset at threshold). Still, the Markov chain formalism (and its extension to infinite memory) can be used. In particular, we exhibit an example where transition probabilities can be explicitly computed and directly related to the dynamics.

We consider a network of N neurons, labeled by the index $k = 1 \dots N$. We define a spike variable $\omega_k(n) = 1$ if neuron k has emitted a spike in the time interval $[n\delta, (n+1)\delta]$, and $\omega_k(n) = 0$ otherwise. We denote by $\omega(n) := [\omega_k(n)]_{k=1}^N$ the spike-state of the entire network at time n , which we call a spiking pattern. We denote by $\mathcal{A} = \{0, 1\}^N$, the state space of spiking patterns in a network of N neurons; a spike block denoted by ω_m^n , $n \geq m$, is the sequence of spike patterns $\omega(m), \omega(m+1), \dots, \omega(n)$; blocks are elements of the product set \mathcal{A}^{n-m} also denoted \mathcal{A}_m^n in the text. We use this last notation because we consider processes with infinite memory ($m \rightarrow -\infty$) and we want to have an explicit notation $\mathcal{A}_{-\infty}^n$ for the corresponding set of events. The time-range (or "range") of a block ω_m^n is $n - m + 1$, the number of time steps from m to n . We call a spike train an infinite sequence of spikes both in the past and in the future. The set of spike trains is thus $\Omega \equiv \mathcal{A}^{\mathbb{Z}}$. To simplify notations we note a spike train $\omega \in \Omega$. The shift operator $\sigma : \Omega \rightarrow \Omega$ is $\sigma\omega = \omega'$, with $\omega'(n) = \omega(n+1)$. This allow us to go one step forward in time along the spike train ω .

We note $\mathcal{F}_{\leq n}$ the set of measurable events (filtration) before time n and \mathcal{F} the filtration on Ω . $\mathcal{P}(\Omega, \mathcal{F})$ is the set of probability measures on (Ω, \mathcal{F}) .

We use the notion of (spike) observable. This is a function $f : \Omega \rightarrow \mathbb{R}$ that associates a real number to a spike-train. We say that the observable $f : \Omega \rightarrow \mathbb{R}$ has range $R = D + 1$ if $f(\omega) \equiv f(\omega_0^D)$. It follows from the Hammersley–Clifford theorem [35,45] that any range- R observable can be written in the form:

$$f(\omega) = \sum_l f_l m_l(\omega), \quad (6)$$

where f_l are real numbers, the coefficients of the decomposition of f in the finite space of range R -observables. The functions m_l spanning this space are called monomials [8]. They have the form:

$$m_l(\omega) = \prod_{k=1}^n \omega_{i_k}(t_k).$$

where $i_k = 1 \dots N$ is a neuron index, and $t_k = 0 \dots D$. Thus, $m_l(\omega) = 1$ if and only if, in the spike train ω , neuron i_1 spikes at time t_1, \dots , neuron i_k spikes at time t_k . Otherwise, $m_l(\omega) = 0$. The number n is the degree of the monomial; degree one monomials have the

form $\omega_{i_1}(t_1)$, degree two monomials have the form $\omega_{i_1}(t_1)\omega_{i_2}(t_2)$, and so on. Thus, monomials are similar to what physicists call (spike) interactions; in our case these interactions involve a time delay between spikes. There are $L = 2^{NR}$ monomials of N neurons and range R and one can index each of them by an integer l in one-to-one correspondence with the set of pairs (i_k, t_k) (see Equation (A2) in the appendix). The advantage of monomial representation is to focus on spike events, which is natural for spiking neuronal dynamics.

We now introduce time-dependent observables. These are functions $f(t, \omega)$ depending on time t (continuous or discrete) and on the spike train ω . The notation $f(t, \omega)$ stands here for $f(t, \omega_{-\infty}^{[t]})$ where $[t]$ is the integer part of t : the function depends on the spike train ω via spikes preceding the current time t . This is an implementation of causality. A range- R time dependent observable is a function $f(t, \omega) \equiv f(t, \omega_{[t]-D}^{[t]})$. The decomposition (6) also holds for a time dependent range- R observable

$$f(t, \omega) = \sum_l f_l(t) m_l(\sigma^{[t]}\omega),$$

where $f_l(t)$ are now functions of time, and $\sigma^{[t]}$ is the $[t]$ -iterate of the time shift operator σ .

2.3. Homogeneous Markov Chains and Gibbs Distributions

A “natural” way to characterize the statistics of observed spike trains is to associate them to a Markov chain with transition probabilities $P_n[\omega(n) | \omega_{n-D}^{n-1}]$, where the index n in P_n indicates that the transition probabilities depend on time n . This approach is “natural” because it captures causality by conditioning on the past spikes. We call D the memory depth of the chain and set $R = D + 1$.

2.3.1. Invariant Probability

Let us start the discussion when transition probabilities are independent of time (homogeneous Markov chain). In this case, we drop the index n in the transition probabilities, $P[\omega(n) | \omega_{n-D}^{n-1}]$. Assuming that all transition probabilities are strictly positive, it follows from the Perron-Frobenius theorem [46,47] that the Markov chain has a unique invariant probability p on \mathcal{A}^D . From the Chapman–Kolmogorov equation [46] one constructs, from p and transition probabilities, a probability measure μ on $\mathcal{P}(\Omega, \mathcal{F})$. where:

$$\mu[\omega_m^n] = \prod_{l=m+D}^n P[\omega(l) | \omega_{l-D}^{l-1}] p[\omega_m^{m+D-1}], \quad \forall m < n \in \mathbb{Z}. \quad (7)$$

As we now discuss, there is a natural correspondence between μ and exponential distributions of the form (4) (Gibbs distributions).

2.3.2. Transfer Matrix

Let us now consider a range- R observable:

$$H(\omega) = \sum_l h_l m_l(\omega), \quad (8)$$

where $h_l > C > -\infty$. Any block ω_0^D of range $R = D + 1$ can be viewed as a transition from a block $\omega^{(u)} = \omega_0^{D-1}$ to the block $\omega^{(u')} = \omega_1^D$. We write $\omega_0^D \sim \omega^{(u)}\omega^{(u')}$. By extension, for two blocks $\omega^{(u)}, \omega^{(u')}$ of range $D \geq 1$ we say that the transition $\omega^{(u)} \rightarrow \omega^{(u')}$ is legal if there is a block $\omega_0^D \sim \omega^{(u)}\omega^{(u')}$. On this basis, one can construct a transfer matrix with positive entries:

$$\mathcal{L}_{\omega^{(u)}, \omega^{(u')}} = \begin{cases} e^{H(\omega_0^D)}, & \text{if } \omega_0^D \sim \omega^{(u)}\omega^{(u')}; \\ 0, & \text{otherwise.} \end{cases}$$

It follows from the Perron-Frobenius theorem [46,47] that \mathcal{L} has a unique real positive eigenvalue s , strictly larger in modulus than the other eigenvalues, and with positive right eigenvector $\mathcal{L}R = sR$, and left eigenvector $L\mathcal{L} = sL$. Moreover, the range- R observable:

$$\phi(\omega_0^D) = H(\omega_0^D) - \log R(\omega_0^{D-1}) + \log R(\omega_1^D) - \log s$$

defines an homogeneous Markov chain [48] with transition probabilities $P[\omega(D) | \omega_0^{D-1}] = e^{\phi(\omega_0^D)}$.

2.3.3. Invariant Probability and Gibbs Distribution

The unique invariant probability of this Markov chain is:

$$p(\omega_0^{D-1}) = R(\omega_0^{D-1})L(\omega_0^{D-1}).$$

Using the Chapman–Kolmogorov Equation (7) one extends p to a probability μ on Ω , where, for $m + n > D$:

$$\mu[\omega_m^n | \omega_{m-D-1}^{m-1}] = \frac{e^{\sum_{l=m-D}^{n-D} H(\omega_l^{l+D})} R(\omega_{n-D+1}^n) L(\omega_{m-D}^{m-1})}{s^{n-m+1}}. \quad (9)$$

This emphasizes the Markovian nature of the process since the conditioning has a finite time horizon of depth D .

It follows therefore that the probability of observing a spike block ω_m^n , given a certain past ω_{m-D-1}^{m-1} is proportional to $e^{\sum_{l=m-D}^{n-D} H(\omega_l^{l+D})}$. If H is formally interpreted as an energy then $\sum_{l=m-D}^{n-D} H(\omega_l^{l+D})$ is the energy of the block ω_m^n . Note that, in contrast to (4) we have removed the $-$ sign which has no reason to be present in this context, and is a source of nuisance when doing computations.

This establishes a first relationship with Gibbs distributions of the form (3), with a strong difference though. More precisely one has, $\exists A, B > 0$ such that, for any block ω_0^n ,

$$A \leq \frac{\mu[\omega_0^n]}{e^{-(n-D+1)\mathcal{P}(\mathcal{H})} e^{-\sum_{k=0}^{n-D} \mathcal{H}(\omega_k^{k+D})}} \leq B,$$

which defines, in ergodic theory, a Gibbs measure in the sense of Bowen [49]. Whereas we assumed (3) to hold on a finite set of states characterizing the system at a given time, here ω is a trajectory of the system describing its time evolution. In addition, the probability of the block ω_m^n is conditioned upon the past, which, in statistical physics would correspond to determine the probability of a block of binary variables (say spins) ω_m^n with left boundary conditions ω_{m-D-1}^{m-1} . This analogy is further developed in the next section.

In the case of a Markov chain, the entropy (5) extends to:

$$S(\mu) = - \sum_{\omega_0^D} p(\omega_0^{D-1}) P[\omega(D) | \omega_0^{D-1}] \log P[\omega(D) | \omega_0^{D-1}],$$

where we have dropped the Boltzmann constant as it plays no role here. Then, it can be shown that μ satisfies a variational principle [48,49]; it maximizes $S[\nu] + E_\nu[H]$, where ν is an invariant probability on $\mathcal{P}(\Omega, \mathcal{F})$. If $E_\nu[H]$ is fixed this amounts to maximizing the entropy under the constraint that the average energy $E_\nu[H]$ is fixed. Finally, the supremum $\mathcal{F} = S[\nu] + E_\nu[H]$ corresponds to the free energy; the generating function of cumulants.

We have therefore shown that a potential of the form (8) is associated with a homogeneous Markov chain where the invariant probability, extends the notion of Gibbs distribution, introduced in Section 2.1, to systems with memory where the probability to be in a state depends on a finite history. The extension to infinite history is made in the next

section. As an important remark, the detailed balance condition is not required to define Gibbs distributions from Markov chains.

2.4. Chains with Infinite Memory and Gibbs Distributions

In the previous section we have made two important assumptions: (i) memory is bounded (finite memory depth D); (ii) the correspondence between Markov chains and Gibbs form was established for homogeneous Markov chains.

However, when considering neural networks, the memory may neither constant nor bounded: consider for example an Integrate and Fire model where the memory goes back to the last time in the past when the neuron has fired; in general, it is not possible to bound this time. So, the most general formalism is to consider chains with unbounded memory [50–52]. Of course, as we discuss below, Markovian approximations are possible and useful. Still, one needs to properly control these approximations. In addition, we want to consider here the case of a system submitted to a time-dependent stimulus, where the dynamic is not time-translation invariant.

Thus, we are now considering a family of transition probabilities of the form $P_n[\omega(n) | \omega_{-\infty}^{n-1}]$, which represent the probability, that at time n , one observes the spiking pattern $\omega(n)$ given the (unbounded) network spike history. Such a non-Markovian stochastic process is known as a “chain with complete connections” or a “chain with unbounded memory” [30] defined in more detail here. This section follows very closely from [53].

Definition 1. A system of transition probabilities is a family $\{P_n\}_{n \in \mathbb{Z}}$ of functions with $P_n[\cdot | \cdot] : \mathcal{A} \times \mathcal{A}_{-\infty}^{n-1} \rightarrow [0, 1]$ such that the following conditions hold for every $n \in \mathbb{Z}$:

- (a) For every $\omega(n) \in \mathcal{A}$ the function $P_n[\omega(n) | \cdot]$ is measurable with respect to $\mathcal{F}_{\leq n-1}$.
- (b) For every $\omega_{-\infty}^{n-1} \in \mathcal{A}_{-\infty}^{n-1}$,

$$\sum_{\omega(n) \in \mathcal{A}} P_n[\omega(n) | \omega_{-\infty}^{n-1}] = 1.$$

Definition 2. A probability measure μ in $\mathcal{P}(\Omega, \mathcal{F})$ is consistent with a system of transition probabilities $\{P_n\}_{n \in \mathbb{Z}}$ if:

$$\int h(\omega_{-\infty}^n) \mu(d\omega) = \int \sum_{\omega(n) \in \mathcal{A}} h(\omega_{-\infty}^{n-1} \omega(n)) P_n[\omega(n) | \omega_{-\infty}^{n-1}] \mu(d\omega), \quad (10)$$

for all $n \in \mathbb{Z}$ and all $\mathcal{F}_{\leq n}$ -measurable functions h . The probability measure μ , when it exists, is called a chain with complete connections consistent with the system of transition probabilities $\{P_n\}_{n \in \mathbb{Z}}$. It is possible that multiple measures are consistent with the same system of transition probabilities.

We now give conditions ensuring the existence of a probability measure consistent with the system of transition probabilities [53].

Definition 3. A system of transition probabilities is non-null on Ω if, for all $n \in \mathbb{Z}$ and all $\omega_{-\infty}^n \in \mathcal{A}_{-\infty}^n$:

$$P[\omega(n) | \omega_{-\infty}^{n-1}] > 0$$

We note, for $n \in \mathbb{Z}$, $m \geq 0$, and r integer:

$$\omega \stackrel{m,n}{=} \omega', \text{ if } \omega(r) = \omega'(r), \forall r \in \{n-m, \dots, n\}.$$

Definition 4. Let m be a positive integer. The m -variation of $P_n[\omega(n) | \cdot]$ is:

$$\text{var}_m[P_n[\omega(n) | \cdot]] = \sup \left\{ |P_n[\omega(n) | \omega_{-\infty}^{n-1}] - P_n[\omega(n) | \omega_{-\infty}^{m-1}]| : \omega \stackrel{m,n}{=} \omega' \right\} \quad (11)$$

Definition 5. The function $P_n[\omega(n) | \cdot]$ is continuous if $\text{var}_m[P_n[\omega(n) | \cdot]] \rightarrow 0$ as $m \rightarrow +\infty$.

The intuitive meaning of continuity is the following. The quantity $\text{var}_m[P_n[\omega(n) | \cdot]]$ corresponds to the maximum variation one can observe on the probability of the spike state at time n , given that the history is fixed up to time $n - m$. Thus, continuity implies that this variation tends to zero as m tends to infinity: the further in the past that the spike sequence is fixed, the smaller the probability that the past influences the present.

The following result holds (see [53]):

Theorem 1. A system of continuous transition probabilities on a compact space has at least one probability measure consistent with it.

Uniqueness requires additional technical assumptions [53]. These conditions hold in the discrete time integrate and fire model [34] considered in Section 4.

Let us now elaborate on the link with Gibbs distributions. First, we define $\phi(n, \omega) : \mathbb{Z} \times \Omega \rightarrow \mathbb{R}$ by:

$$\phi(n, \omega) \equiv \log P[\omega(n) | \omega_{-\infty}^{n-1}], \quad (12)$$

and:

$$\Phi(m, n, \omega) = \sum_{r=m}^n \phi(r, \omega). \quad (13)$$

Then:

$$P[\omega_m^n | \omega_{-\infty}^{m-1}] = e^{\Phi(m, n, \omega)} = e^{\sum_{r=m}^n \phi(r, \omega)} \quad (14)$$

and:

$$\mu[\omega_m^n] = \int_{\mathcal{A}_{-\infty}^{m-1}} e^{\Phi(m, n, \omega)} \mu(d\omega). \quad (15)$$

These equations emphasize the connection with Gibbs distributions in statistical physics where ϕ acts as an “energy” [53,54]. From now on we use the term “potential” instead. The correspondence in our case is to consider “time” as a 1-dimensional “lattice” and the “boundary conditions” as the past $\omega_{-\infty}^{m-1}$ of the stochastic process. In contrast to statistical physics, and because the potential is defined via transition probabilities, the normalization factor (partition function) is equal to 1. For this reason, we call ϕ a normalized Gibbs potential.

Equations (14) and (15) are similar to (9) with an essential difference: the memory is now infinite, and the potential ϕ has an infinite range. As it is well-known in statistical physics [38,39], infinite range potentials require specific conditions to be associated with a unique Gibbs distribution. There is a mathematically well-founded correspondence between chains with complete connections and Gibbs distributions [38,39,53]. However, while chains with complete connections define transition probabilities where the present is conditioned upon the past, Gibbs distributions allow conditioning “upon the future” as well. More generally, Gibbs distributions in statistical physics extend to probability distributions on \mathbb{Z}^d where the probability (3) to observe a certain configuration of spins in a restricted region of space is constrained by the configuration at the boundaries of this region. They are therefore defined in terms of specifications [38,39], which determine finite-volume conditional probabilities when the exterior of the volume is known. In one spatial dimension ($d = 1$), identifying \mathbb{Z} with a time axis, this corresponds to conditioning both in the past and in the future. In contrast, families of transition probabilities with an exponential continuity rate define the so-called left-interval specifications (LIS) [53,55].

This leads to not equivalent notions of “Gibbsianness” [56]. We shall not develop on these distinctions here and call Gibbs distribution a chain with the complete connection.

3. Linear Response for Neuronal Networks with Unbounded Memory

We consider a neural system where spike statistics is characterized by a time-translation invariant Gibbs distribution (chain with unbounded memory) $\mu^{(sp)}$ where “sp” stands for “spontaneous”. That is, we suppose that, in the absence of a stimulus, the spontaneous dynamics is stationary. We assume that a stimulus $S(t)$ is applied from time $t = t_0$, and that conditions of existence and uniqueness of a chain with complete connection μ are fulfilled in the presence of the stimulus (an example is given in the next section). We note $n_0 = [t_0]$. For times anterior to n_0 , μ identifies with $\mu^{(sp)}$, that is, for any $m < n \leq n_0$, for any block ω_m^n , $\mu[\omega_m^n] = \mu^{(sp)}[\omega_m^n]$. In contrast, for $n > n_0$ spike statistics is modified. Consider a range- R observable $f(t, \omega)$ then $E_\mu[f(t, \cdot)] \stackrel{\text{def}}{=} E_{\mu^{(sp)}}[f(t, \cdot)] + \delta[f(t, \cdot)]$ where $\delta[f(t, \cdot)] = 0$ for $t < t_0$ and $\delta[f(t, \cdot)] \neq 0$ for $t \geq t_0$.

The goal is to establish an explicit (formal) equation for $\delta[f(t, \cdot)]$, as a function of the stimulus. This is done via a Volterra-like expansion in powers of the stimulus, cut to the first order so as to obtain a linear response in terms of a convolution between the stimulus and a convolution kernel K_f , depending on f , $\delta^{(1)}[f(t, \cdot)] = [K_f * S](t)$. This way, we obtain a relationship between the proportionality coefficient K_f in the linear response, and specific correlation functions computed at equilibrium (spontaneous activity). This provides a Kubo relation holding in the case of neuronal networks with unbounded memory and for an arbitrary range- R observable f . In contrast to Volterra expansion, our formalism allows to explicit the dependence of K_f in the neuronal network characteristics (parameters fixing the individual neuron dynamics and connectivity-synaptic weights). An example is provided in the next section.

3.1. First Order Expansion

We assume that the statistics of spikes is described by time-dependent chains with unbounded memory, with potential $\phi(n, \omega)$. We note $\delta\phi(n, \omega) = \phi(n, \omega) - \phi^{(sp)}(\omega)$. We define likewise $\Phi(n, \omega) = \Phi^{(sp)}(\omega) + \delta\Phi(n, \omega)$ using (13).

From the definition (12), $e^{\phi^{(sp)}}$ corresponds to the family of transition probabilities $\{P^{(sp)}\}$ defining the Gibbs distribution $\mu^{(sp)}$ in the spontaneous regime, whereas e^ϕ corresponds to the family of transition probabilities $\{P\}$ defining the Gibbs distribution μ in time-dependent stimuli-evoked regime. For $n > n_0$, we have:

$$e^{\Phi(n_0+1, n, \omega)} = e^{\Phi^{(sp)}(n_0+1, n, \omega) + \delta\Phi(n_0+1, n, \omega)} = e^{\Phi^{(sp)}(n_0+1, n, \omega)} \left[1 + \sum_{p=1}^{+\infty} \frac{\delta\Phi(n_0+1, n, \omega)^p}{p!} \right],$$

which, from Equation (13), gives:

$$P[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}] = P^{(sp)}[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}] \left[1 + \sum_{p=1}^{+\infty} \frac{\delta\Phi(n_0+1, n, \omega)^p}{p!} \right].$$

Taking the first-order approximation of the exponential, we obtain:

$$e^{\Phi(n_0+1, n, \omega)} \sim e^{\Phi^{(sp)}(n_0+1, n, \omega)} [1 + \delta\Phi(n_0+1, n, \omega)].$$

However, while $\Phi(n_0+1, n, \omega)$ and $\Phi^{(sp)}(n_0+1, n, \omega)$ are normalized potentials, i.e., the log of a conditional probability, the first order approximation of $e^{\Phi^{(sp)}(n_0+1, n, \omega)}$

$[1 + \delta\Phi(n_0 + 1, n, \omega)]$ is not. Normalization is obtained formally by introducing the partition function:

$$Z[\omega_{-\infty}^{n_0}] = \sum_{\omega_{n_0+1}^n} e^{\Phi^{(sp)}(n_0+1, n, \omega)} [1 + \delta\Phi(n_0 + 1, n, \omega)],$$

constrained by the past sequence $\omega_{-\infty}^{n_0}$, so that the quantity

$$P^{(1)}[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}] \equiv \frac{e^{\Phi^{(sp)}(n_0+1, n, \omega)}}{Z[\omega_{-\infty}^{n_0}]} [1 + \delta\Phi(n_0 + 1, n, \omega)],$$

is the first order approximation of $P[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}]$.

Setting:

$$Z^{(sp)}[\omega_{-\infty}^{n_0}] = \sum_{\omega_{n_0+1}^n} e^{\Phi^{(sp)}(n_0+1, n, \omega)},$$

we have, to first order:

$$\frac{1}{Z[\omega_{-\infty}^{n_0}]} = \frac{1}{Z^{(sp)}[\omega_{-\infty}^{n_0}]} \left[1 - \sum_{\omega_{n_0+1}^n} \frac{e^{\Phi^{(sp)}(n_0+1, n, \omega)}}{Z^{(sp)}[\omega_{-\infty}^{n_0}]} \delta\Phi(n_0 + 1, n, \omega) \right].$$

However, as $\Phi^{(sp)}$ is the log of a conditional probability, $Z^{(sp)}[\omega_{-\infty}^{n_0}] = 1$. So, finally, we obtain, to first order:

$$P^{(1)}[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}] \sim P^{(sp)}[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}] \left[1 + \delta\Phi(n_0 + 1, n, \omega) - E^{(sp)}[\delta\Phi(n_0 + 1, n, \cdot) | \omega_{-\infty}^{n_0}] \right],$$

where $E^{(sp)}[\cdot]$ denotes the expectation with respect to $\mu^{(sp)}$. We use $E^{(sp)}[\cdot]$ instead of $E_{\mu^{(sp)}}[\cdot]$ to alleviate notations.

3.2. Time Dependent Average of an Observable

We now consider a time-dependent observable f with finite range $R \equiv R_f$. We assume $t - t_0 > R_f$. We set $R_f = D_f + 1$. Setting $n = [t]$ we note $E_n[f(t, \cdot)] = \int f(t, \omega) \mu(d\omega)$. Here, a note of explanation is necessary. Functions $f(t, \omega)$ are random functions, where the randomness comes from ω . So, the law of $f(t, \omega)$ is determined by the probability μ . $E_n[f(t, \cdot)]$ is the average of the continuous time dependent observable $f(t, \omega)$, averaged over the discrete time spike train ω , up to the discrete time $n = [t]$ (by definition $f(t, \cdot)$ does not depend on spike events occurring at times posterior to n). Note that this average cannot be defined by an ergodic time average procedure as, here, the probability is non stationary (see Section 5 for a numerical implementation).

Because f has finite range R_f we may write:

$$E_n[f(t, \cdot)] = \sum_{\omega_{n-D_f}^n} f(t, \omega_{n-D_f}^n) \mu[\omega_{n-D_f}^n] = \sum_{\omega_{n_0+1}^n} f(t, \omega_{n-D_f}^n) \mu[\omega_{n_0+1}^n].$$

The last equality holds because $f(t, \omega)$ is independent of $\omega_{-\infty}^{n_0}$. Thus, using (15):

$$\begin{aligned} E_n[f(t, \cdot)] &= \sum_{\omega_{n_0+1}^n} f(t, \omega_{n-D_f}^n) \int_{\mathcal{A}_{-\infty}^{n_0}} P[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}] \mu(d\omega) \\ &= \sum_{\omega_{n_0+1}^n} f(t, \omega_{n-D_f}^n) \int_{\mathcal{A}_{-\infty}^{n_0}} P[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}] \mu^{(sp)}(d\omega), \end{aligned}$$

where the last equation holds because, on $\mathcal{F}_{\leq n_0}$, $\mu = \mu^{(sp)}$.

Thus, replacing $P[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}]$ by $P^{(1)}[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}]$, we obtain, up to first order:

$$\begin{aligned} E_n[f(t, \cdot)] &\sim \sum_{\omega_{n_0+1}^n} f(t, \omega_{n-D_f}^n) \int_{\mathcal{A}_{-\infty}^{n_0}} P^{(sp)}[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}] \mu^{(sp)}(d\omega) \\ &+ \sum_{\omega_{n_0+1}^n} f(t, \omega_{n-D_f}^n) \int_{\mathcal{A}_{-\infty}^{n_0}} P^{(sp)}[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}] \delta\Phi(n_0 + 1, n, \omega) \mu^{(sp)}(d\omega) \\ &- \sum_{\omega_{n_0+1}^n} f(t, \omega_{n-D_f}^n) \int_{\mathcal{A}_{-\infty}^{n_0}} P^{(sp)}[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}] E^{(sp)}[\delta\Phi(n_0 + 1, n, \cdot) | \omega_{-\infty}^{n_0}] \mu^{(sp)}(d\omega). \end{aligned}$$

The first term is $E^{(sp)}[f(t, \cdot)]$ from (15). The second term is:

$$\sum_{\omega_{n_0+1}^n} \int_{\mathcal{A}_{-\infty}^{n_0}} f(t, \omega) \delta\Phi(n_0 + 1, n, \omega) P[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}] \mu^{(sp)}(d\omega) = E^{(sp)}[f(t, \cdot) \delta\Phi(n_0 + 1, n, \cdot)]$$

from the consistency property (10), and because by assumption $(n - n_0 > R_f)$, $f(t, \omega)$ does not depend on $\omega_{-\infty}^{n_0}$.

For the third term:

$$\begin{aligned} \sum_{\omega_{n_0+1}^n} f(t, \omega_{n-D_f}^n) \int_{\mathcal{A}_{-\infty}^{n_0}} E^{(sp)}[\delta\Phi(n_0 + 1, n, \omega) | \omega_{-\infty}^{n_0}] P^{(sp)}[\omega_{n_0+1}^n | \omega_{-\infty}^{n_0}] \mu^{(sp)}(d\omega) \\ = E^{(sp)}[f(t, \cdot) E^{(sp)}[\delta\Phi(n_0 + 1, n, \omega) | \omega_{-\infty}^{n_0}]]. \end{aligned}$$

However, by assumption $f(t, \omega_{n-D_f}^n)$ does not depend on $\omega_{-\infty}^{n_0}$ ($n - n_0 > D_f$), whereas by definition of the conditional expectation $E^{(sp)}[\delta\Phi(n_0 + 1, n, \omega) | \omega_{-\infty}^{n_0}]$ is the projection on the sigma-algebra $\mathcal{F}_{\leq n_0}$. As a consequence, we have:

$$\begin{aligned} E^{(sp)}[f(t, \cdot) E^{(sp)}[\delta\Phi(n_0 + 1, n, \cdot) | \omega_{-\infty}^{n_0}]] &= E^{(sp)}[f(t, \cdot)] E^{(sp)}[E^{(sp)}[\delta\Phi(n_0 + 1, n, \cdot) | \omega_{-\infty}^{n_0}]] \\ &= E^{(sp)}[f(t, \cdot)] E^{(sp)}[\delta\Phi(n_0 + 1, n, \cdot)]. \end{aligned}$$

Summing up, we have, using (13):

$$E_n[f(t, \cdot)] = E^{(sp)}[f(t, \cdot)] + \sum_{r=n_0+1}^{n=[t]} \left(E^{(sp)}[f(t, \cdot) \delta\phi(r, \cdot)] - E^{(sp)}[f(t, \cdot)] E^{(sp)}[\delta\phi(r, \cdot)] \right),$$

Using the correlation function:

$$\mathcal{C}^{(sp)}[f(t, \cdot), g(t', \cdot)] \stackrel{\text{def}}{=} E^{(sp)}[f(t, \cdot) g(t', \cdot)] - E^{(sp)}[f(t, \cdot)] E^{(sp)}[g(t', \cdot)]$$

we obtain

$$\delta^{(1)}[f(t, \cdot)] = \sum_{r=n_0+1}^{n=[t]} \mathcal{C}^{(sp)}[f(t, \cdot), \delta\phi(r, \cdot)]. \quad (16)$$

This equation expresses that the time-dependent variation in the average of an observable f is expressed, to the first order, as a time series of correlation functions, between f and the time-dependent variation of the normalized potential, computed with respect to the equilibrium distribution. This is our main result.

It is similar to the fluctuation-dissipation theorem in statistical physics [32,33]. Here, it holds for Gibbs distributions with infinite range potential $\phi(t, \omega)$. A crucial point is the convergence of the series when the initial time of perturbation, n_0 tends to $-\infty$. This

holds if correlations $\mathcal{C}^{(sp)}[f(t, \cdot), \delta\phi(r, \cdot)]$ decay sufficiently fast, typically, exponentially. We consider only this case in this article.

One of the advantages of this relationship is that averages are taken with respect to $\mu^{(sp)}$. In the case of experimental data, these averages can be approximated by empirical averages on spontaneous activity. The monomial decomposition of (16) can be found in the Appendix B.

4. An Example: Linear Response in a Conductance Based Integrate and Fire Model

4.1. Discrete Time Integrate and Fire Network Model

The Leaky Integrate and Fire model (LIF) consider a point neuron k (without spatial extension), with membrane potential V_k , membrane capacity C_k , resistance R , submitted to a current $I_k(t)$. Call θ the spiking threshold. The sub-threshold dynamics is:

$$C_k \frac{dV_k}{dt} + \frac{1}{R} V_k = I_k(t), \quad \text{if } V_k(t) < \theta. \quad (17)$$

If there is a time t_k such that the membrane potential of neuron k reaches the firing threshold, $V_k(t_k) \geq \theta$, the neuron k fires an action potential, i.e., it emits a spike and the membrane potential of neuron k is reset to a fixed reset value V_{res} instantaneously. The neuron's membrane potential remains at this value during a time denoted by Δ called "refractory period", i.e., $V_k(t') = V_{res}$, $t' \in [t_k, t_k + \Delta]$.

To illustrate our results we use a "simple" model corresponding to a discrete-time LIF network [34]. The main reason to choose this model is to facilitate numerical simulations as handling continuous-time dynamics with spikes event, although manageable mathematically [57] is difficult to handle numerically [58]. To further simplify the analysis we fixing the sampling time $dt = 1$, the capacitance $C_k = 1$, $\tau = RC$, $\gamma = 1 - 1/\tau \in [0, 1[$ is called the "leak rate". The evolution of the membrane potential V_k is ruled by the following sub-threshold equation:

$$V_k(n+1) = \gamma V_k(n) + I_k(t).$$

Note that by setting $C_k = 1$ we are not considering units of measurements anymore.

We now explicitly consider an interconnected network of neurons. The network of synaptic connectivity is represented by a matrix of components W_{kj} which can be positive or negative to characterize inhibition or excitation respectively. We consider random fluctuations by adding a standard Gaussian additive noise $\xi_k(n)$ controlled by the amplitude σ_B . We also consider a constant stimulus I_0 and a time dependent stimulus $S_k(t)$. Thus $I_k(t) = \sum_j W_{kj} \omega_j(n) + I_0 + S_k(t) + \sigma_B \xi_k(n)$, and thus our equation finally reads:

$$V_k(n+1) = \gamma V_k(n) + \sum_j W_{kj} \omega_j(n) + I_0 + S_k(t) + \sigma_B \xi_k(n), \quad \text{if } V_k(n) < \theta. \quad (18)$$

Note that γ , the decay term, is related to the leak characteristic time (18) by:

$$\tau = \frac{1}{1 - \gamma}.$$

The condition $\gamma < 1$ define the exponential decay in the spike history dependence via the characteristic time:

$$\tau_\gamma = - \left[\frac{1}{\log \gamma} \right]. \quad (19)$$

This characteristic time can be interpreted as follows. Integrating Equation (18) up to the last time where voltage was reset in the past, $\tau_k(n, \omega)$ gives the following:

$$V_k(n+1, \omega) = \sum_{j=1}^N W_{kj} \eta_{kj}(n, \omega) + I_0 \frac{1 - \gamma^{n+1 - \tau_k(n, \omega)}}{1 - \gamma} + \sum_{l=\tau_k(n, \omega)}^n \gamma^{n-l} S_k(l) + \sigma_B \sum_{l=\tau_k(n, \omega)}^n \gamma^{n-l} \xi_k(l), \quad (20)$$

where:

$$\eta_{kj}(n, \omega) = \sum_{l=\tau_k(n, \omega)}^n \gamma^{n-l} \omega_j(l),$$

The first term on the right-hand side of (20) corresponds to the synaptic contribution. The second corresponds to the integration of the constant term I_0 , used to fix the baseline activity. The third term corresponds to the integration of a time-dependent stimulus $S_k(t)$ and the fourth to the integrated noise term with intensity σ_B .

For each $n \in \mathbb{Z}$, conditionally to $\omega_{-\infty}^{n-1}$, $V_k(n)$ is Gaussian random variable. It can be decomposed in the following way:

$$V_k(n, \omega) = V_k^{(syn)}(n, \omega) + V_k^{(I)}(n, \omega) + V_k^{(S)}(n, \omega) + V_k^{(noise)}(n, \omega)$$

We now consider the “spontaneous” voltage $V_k^{(sp)}(n, \omega)$ and the evoked response due to the external time dependent stimulus $V_k^{(S)}(n, \omega)$.

$$\begin{aligned} V_k^{(sp)}(n, \omega) &= V_k^{(syn)}(n, \omega) + V_k^{(I)}(n, \omega) + V_k^{(noise)}(n, \omega) \\ &= \sum_{j=1}^N W_{kj} \eta_{kj}(n, \omega) + I_0 \frac{1 - \gamma^{n+1-\tau_k(n, \omega)}}{1 - \gamma} + \sigma_B \sum_{l=\tau_k(n, \omega)}^n \gamma^{n-l} \xi_k(l) \end{aligned} \quad (21)$$

$$V_k^{(S)}(n, \omega) = \sum_{l=\tau_k(n, \omega)}^n \gamma^{n-l} S_k(l)$$

4.2. Transition Probabilities of the Discrete Time LIF Model

In the limit of small σ_B , the family of transition probabilities $P_n[\omega(n) | \omega_{-\infty}^{n-1}]$ can be written explicitly in terms of the parameters of the spiking neuronal network model. The discrete-time LIF model is conditionally independent i.e., it factorizes over neurons once the spike history has been fixed [34]. The same property is held for the continuous-time version of this model [59]. However, a more complete version of this model also includes electric synapses [57]. In that case, the conditional independence is lost.

$$P_n[\omega(n) | \omega_{-\infty}^{n-1}] = \prod_{k=1}^N P_n[\omega_k(n) | \omega_{-\infty}^{n-1}],$$

where:

$$P_n[\omega_k(n) | \omega_{-\infty}^{n-1}] = \omega_k(n) \Pi(X_k(n-1, \omega)) + (1 - \omega_k(n)) (1 - \Pi(X_k(n-1, \omega))), \quad (22)$$

with:

$$\Pi(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-\frac{u^2}{2}} du,$$

and:

$$X_k(n-1, \omega) \stackrel{\text{def}}{=} \frac{\theta - V_k(n-1, \omega)}{\sigma_k(n-1, \omega)}, \quad (23)$$

where,

$$\sigma_k^2(n-1, \omega) = \frac{1 - \gamma^{2(n-\tau_k(n-1, \omega))}}{1 - \gamma^2}$$

corresponds to the variance of the noise integrated up to time $n-1$.

Combining (12) and (22), we obtain the normalized potential for the discrete time LIF model.

$$\phi(n, \omega) = \sum_{k=1}^N \phi_k(n, \omega),$$

where

$$\phi_k(n, \omega) = \omega_k(n) \log \Pi(X_k(n-1, \omega)) + (1 - \omega_k(n)) \log(1 - \Pi(X_k(n-1, \omega))), \quad (24)$$

which depends on all of the parameters of the network via the variable $X_k(n-1, \omega)$ (23).

Remark 1. The function $\Pi(x)$ is a sigmoid which tends to 1 when $x \rightarrow -\infty$ and tends to 0 when $x \rightarrow \infty$. This has two consequences:

1. When $X_k(n-1, \omega) \rightarrow +\infty$ (which arises when $V_k^{(sp)}(n-1, \omega) \rightarrow -\infty$), $\Pi(X_k(n-1, \omega)) \rightarrow 0$ so that $\phi_k(n, \omega) \rightarrow -\infty$. This expresses that the probability of having a spike at time n when $X_k(n-1, \omega)$ becomes large (neuron strongly hyper-polarized) tends to 0. The same argument holds mutatis mutandis for the limit $X_k(n-1, \omega) \rightarrow -\infty$.
2. When $|X_k(n-1, \omega)|$ is large (neuron either strongly hyper-polarized or strongly depolarized) the effect of a variation of the membrane potential on the firing probability is negligible. Thus, we study the effect of a perturbation in bounded range for $X_k(n-1, \omega)$:

$$0 < \epsilon < \Pi(X_k(n-1, \omega)) < 1 - \epsilon < 1, \quad (25)$$

uniformly in n, ω . This is ensured by natural assumptions on synaptic weights and on σ_B , the mean-square deviation of the noise, which has to be bounded away from 0.

4.3. Expansion of the Normalized Potential

The normalized potential of the Discrete time LIF model can be separated into: (i) a “spontaneous” part $\phi^{(sp)}(\omega)$, which before time t_0 is independent of the stimuli and time and; (ii) a “perturbation” part $\delta\phi(n, \omega)$ depending on a time-dependent stimuli, which is non-zero from time t_0 . Mathematically this is achieved by adding an extra term to the spontaneous potential after time t_0 .

$$\phi(n, \omega) = \begin{cases} \phi^{(sp)}(\omega) & \text{if } n < [t_0]; \\ \phi^{(sp)}(\omega) + \delta\phi(n, \omega) & \text{if } n \geq [t_0]. \end{cases}$$

Note that, at this stage, this is just a definition of $\delta\phi(n, \omega) = \phi(n, \omega) - \phi^{(sp)}(\omega)$.

From Section 3 this perturbation induces a time-dependent variations on the average of an observable f :

$$\mu_n[f(n, \cdot)] = E^{(sp)}[f(n, \cdot)] + \delta^{(1)}[f(n, \cdot)],$$

If $n \leq t_0$, $\delta^{(1)}[f(n, \cdot)] = 0, \forall f(n, \cdot)$ as $\mu_n = \mu^{(sp)}$. Thus, the term $E^{(sp)}[f(n, \cdot)]$ refers to an average with respect to the unperturbed system and $\delta^{(1)}[f(n, \cdot)]$. As we show, the variation $\delta^{(1)}$ can be explicitly written in terms of the variation on the normalized potential induced by the introduction of the stimulus.

Note that if the external stimuli are switched on at time t_0 , spike statistics are still constrained by the previous spontaneous activity, since transition probabilities have memory. This effect is especially salient in the discrete-time LIF model which has an unbounded memory.

We rewrite (23) in the form:

$$X_k(n-1, \omega) = X_k^{(sp)}(n-1, \omega) + \delta X_k(n-1, \omega),$$

where

$$X_k^{(sp)}(n-1, \omega) = \frac{\theta - V_k^{(sp)}(n-1, \omega)}{\sigma_k(n-1, \omega)}$$

is independent of the stimulus, and:

$$\delta X_k(n-1, \omega) = -\frac{V_k^{(S)}(n-1, \omega)}{\sigma_k(n-1, \omega)} = -\frac{1}{\sigma_k(n-1, \omega)} \sum_{l=\tau_k(n-1, \omega)}^{n-1} \gamma^{n-1-l} S_k(l),$$

where the last equality holds because $S(n) = 0$ for $n < t_0$.

In the next computation we write $X_k^{(sp)}$, δX_k instead of $X_k^{(sp)}(n-1, \omega)$, $\delta X_k(n-1, \omega)$ to alleviate notations. We make a series expansion of $\phi_k(n, \omega)$ at $X_k^{(sp)}$, under the conditions of (25). We have:

$$\begin{aligned} \log \Pi(X_k^{(sp)} + \delta X_k) &= \log \Pi(X_k^{(sp)}) + \sum_{u=1}^{+\infty} \frac{a^{(u)}(X_k^{(sp)})}{u!} (\delta X_k)^u, \\ \log(1 - \Pi(X_k^{(sp)} + \delta X_k)) &= \log(1 - \Pi(X_k^{(sp)})) + \sum_{u=1}^{+\infty} \frac{b^{(u)}(X_k^{(sp)})}{u!} (\delta X_k)^u \end{aligned}$$

where $a^{(u)}$ and $b^{(u)}$ are the u -th derivative of $\log \Pi(x)$ and $\log(1 - \Pi(x))$. In particular:

$$a^{(1)}(x) = \frac{\Pi'(x)}{\Pi(x)}; \quad b^{(1)}(x) = -\frac{\Pi'(x)}{1 - \Pi(x)}.$$

Therefore,

$$\delta \phi_k(n, \omega) = \sum_{u=1}^{+\infty} \delta \phi_k^{(u)}(n, \omega),$$

where:

$$\delta \phi_k^{(u)}(n, \omega) = \mathcal{H}_k^{(u)}(n, \omega) [\delta X_k(n-1, \omega)]^u,$$

with:

$$\mathcal{H}_k^{(u)}(n, \omega) = \frac{1}{u!} \left[\omega_k(n) a^{(u)}(X_k^{(sp)}(n-1, \omega)) + (1 - \omega_k(n)) b^{(u)}(X_k^{(sp)}(n-1, \omega)) \right]. \quad (26)$$

This expansion holds for any value of $X_k^{(sp)}(n-1, \omega)$. However, when this quantity becomes large in absolute value, one has to consider more and more terms in the expansion to approach sufficiently well the function $\delta \phi_k(n, \omega)$. This is a well-known property of the function Π (which can be written in terms of the error function): the Taylor expansion converges very slowly near infinity and other expansions are more efficient (e.g., Bürmann series [60]). Here, we consider the effect of a perturbation in a range where the function Π does not saturate. In addition, we restrict ourselves to cases where the first order of the Taylor expansion is sufficient to characterize the response. This is ensured by conditions of the form (the same holds mutatis mutandis for b):

$$|\delta X_k| \ll (u!)^{\frac{1}{u-1}} \left| \frac{a^{(1)}(X_k^{(sp)})}{a^{(u)}(X_k^{(sp)})} \right|^{\frac{1}{u-1}}; \quad u > 1.$$

Applied to the second order this gives a condition:

$$|\delta X_k(n-1, \omega)| \ll \frac{2}{\left| X_k^{(sp)}(n-1, \omega) + a^{(1)}(X_k^{(sp)}(n-1, \omega)) \right|}, \quad (27)$$

which becomes more and more restrictive as one gets away from $X_k^{(sp)}(n-1, \omega) = 0$ i.e., away from the linear region of the sigmoid Π . Note that the condition $X_k^{(sp)}(n-1, \omega) \sim 0$ corresponds to a voltage close to the firing threshold. We insist that this constraint is not a limitation of our approach, but instead, a limitation of linear response applied in neuronal systems where the response of a neuron is characterized by a saturating function. Away from the linear part of the sigmoid, nonlinear effects dominate the response. In the numerical simulations, we use the mean-field approximation where we replace $\tau_k(r-1, \cdot)$ by its average value $\mu[\tau_k(r-1, \cdot)] = r-1 - \frac{1}{\nu_k}$ where the inverse $\frac{1}{\nu_k}$ of the average firing rate ν_k of neuron k is the mean time between two spikes.

Under these conditions we obtain:

$$\delta\phi_k^{(1)}(r, \omega) = -\frac{\mathcal{H}_k^{(1)}(r, \omega)}{\sigma_k(r-1, \omega)} \sum_{l=r-1-\frac{1}{\nu_k}}^{r-1} \gamma^{r-1-l} S_k(l)$$

where we have replaced n by r in view of (16).

The function $\delta\phi_k^{(1)}(r, \omega)$ is the first order variation of normalized potential when neurons are submitted to a weak time dependent stimulus, under the approximation (24). There are two contributions. The integral includes the effect of the stimulus on the dynamics flow; the term $\mathcal{H}_k^{(1)}(r, \omega)$, given by Equation (26), contains the effect of the network via the terms $a^{(1)}\left(\frac{\theta - V_k^{(sp)}(r-1, \omega)}{\sigma_k(n-1, \omega)}\right)$, $b^{(1)}\left(\frac{\theta - V_k^{(sp)}(r-1, \omega)}{\sigma_k(n-1, \omega)}\right)$ where $V_k^{(sp)}(r-1, \omega)$ is given by Equation (21). Note that the dependence on synaptic weights is non-linear because $a^{(1)}$ and $b^{(1)}$ are non-linear.

In order to study the linear response theory in this model, one has to properly handle numerically:

- (1) The spike train statistics, especially finding a numerical way to perform a suitable averaging, not only for the spontaneous probability measure where time-ergodic average can be used but also for the non-stationary response where ergodicity does not take place;
- (2) The long memory tail in the dynamics;
- (3) Find an illustrative example with a good range of parameters showing convincing, original, non-trivial results while avoiding prohibitive computational times.

It is not evident to us that classical spiking neural network simulators such as BRIAN [61], although quite efficient, could easily handle (1) in conjunction with (2). The model presented below, has been studied both from the mathematical side and the numerical side [8,34]. Additionally, we have designed a simulation tool, PRANAS, devoted to the analysis of population spike train statistics and allowing to properly handle Gibbs distributions from numerical simulations or experimental data, with a specific module dedicated to this model [62]. This software is freely downloadable at <https://team.inria.fr/biovision/pranas-software/> on simple demand. The linear response C++ codes and instructions required to reproduce these numerical results can be found at <https://github.com/bricolab/Linear-response>.

4.4. Linear Response

We apply our main result (16) to the potential of the discrete-time model (24). Under the approximations made in the previous section, we examine two forms of linear response proposed in the paper.

1. The first-order expansion of the potential, (16). In the present context it becomes:

$$\delta\mu^{(1)}[f(n)] \sim -\sum_{k=1}^N \sum_{r=-\infty}^n \sum_{l=r-1-\frac{1}{\nu_k}}^{r-1} \gamma^{r-1-l} \mathcal{C}^{(sp)}[f(n, \cdot), \zeta(r-1, \cdot)] S_k(l),$$

where we have set $\zeta_k(r-1, \omega) = \frac{\mathcal{H}_k^{(1)}(r, \omega)}{\sigma_k(r-1, \omega)}$. In the numerical simulations we use a Markovian approximation with memory depth D such that the sum $\sum_{r=-\infty}^n$ is replaced by $\sum_{r=n-D}^n$. D is typically determined by exponential decay of the terms $\gamma^{l-r+1} \mathcal{C}^{(sp)}[f(n, \cdot), \zeta_k(r-1, \cdot)]$, which is controlled, on one hand, by γ (with a characteristic time $\tau_\gamma = -\left\lceil \frac{1}{\log \gamma} \right\rceil$), and the correlation $\mathcal{C}^{(sp)}[f(n, \cdot), \zeta_k(r-1, \cdot)]$, which is controlled by the spectral gap in the Perron–Frobenius matrix. In this approximation we have:

$$\delta\mu^{(1)}[f(n)] \sim - \sum_{k=1}^N \sum_{r=n-D}^n \sum_{l=r-1-\frac{1}{v_k}}^{r-1} \gamma^{r-1-l} \mathcal{C}^{(sp)}[f(n, \cdot), \zeta_k(r-1, \cdot)] S_k(l).$$

Using the stationarity of $\mu^{(sp)}$ we have $\mathcal{C}^{(sp)}[f(n, \cdot), \zeta(r-1, \cdot)] = \mathcal{C}^{(sp)}[f(n-r+1, \cdot), \zeta_k(0, \cdot)]$ where $r \in [n-D, n]$ so that $m = n-r+1 \in [1, D+1]$. Then, introducing the $N \times D$ matrix $\mathcal{K}^{(1)}$ with entries :

$$\mathcal{K}_{k,m}^{(1)} = \mathcal{C}^{(sp)}[f(m, \cdot), \zeta_k(0, \cdot)],$$

we may write $\delta\mu^{(1)}[f(n)]$ in the form:

$$\delta\mu^{(1)}[f(n)] \sim - \sum_{k=1}^N \sum_{m=1}^{D+1} \sum_{l=0}^{\frac{1}{v_k}} \gamma^l \mathcal{K}_{km}^{(1)} S_k(n-m-l). \quad (28)$$

2. The first order Hammersley–Clifford expansion of $\mathcal{H}_k^{(1)}$ which corresponds to expand $\zeta_k(0, \omega)$ to the lowest order in the Hammersley–Clifford expansion, $\zeta_k(0, \omega) \sim \gamma_k^{(1)} \omega_k(0)$ with:

$$\gamma_k^{(1)} = a^{(1)}(\theta_{L,k}) - b^{(1)}(\theta_{L,k}),$$

with $a^{(1)}(x) = \frac{\Pi'(x)}{\Pi(x)}$, $b^{(1)}(x) = -\frac{\Pi'(x)}{1-\Pi(x)}$. This gives an approximation similar to the fluctuation-dissipation theorem where the linear response is a sum of pairwise correlation functions. In the discrete time model it reads:

$$\delta\mu^{(HC_1)}[f(n)] \sim -\frac{2\sqrt{1-\gamma}}{\sigma_B} \gamma^{(1)} \sum_{k=1}^N \sum_{m=1}^{D+1} \sum_{l=0}^{\frac{1}{v_k}} \gamma^l \mathcal{K}_{km}^{(HC_1)} S_k(n-m-l). \quad (29)$$

where :

$$\mathcal{K}_{k,m}^{(HC_1)} = \mathcal{C}^{(sp)}[f(m, \cdot), \omega_k(0, \cdot)].$$

We have used the superscript “ HC_1 ” to refer to the lowest-order Hammersley–Clifford approximation from (28). The interest of this approximation (and more generally, of the Hammersley–Clifford expansion is that it can be obtained without knowing explicitly the potential ϕ (by a mere fit of the coefficients $\gamma_k^{(1)}$). We expect however (28) to give a better approximation of the linear response than (29). Note that the effective threshold corresponds to $\theta_{L,k} = \frac{\theta - \frac{I_0}{1-\gamma}}{\sqrt{\frac{1}{2(1-\gamma)} \sigma_B}}$, is independent of k in this case.

This explains why $\gamma^{(1)}$ gets out of the sum and lost its index k in (29).

5. Numerical Simulations

5.1. Averaging Method

We need to compute numerically averages with respect to the invariant probability $\mu^{(sp)}$. As $\mu^{(sp)}$ is ergodic it is in principle possible to get them by time averaging. However, we also want to compute averages in the presence of a stimulus, where ergodicity does not take place. In addition, a notation like $E^{(sp)}[f(n, \cdot) g(r, \cdot)]$ appearing all over the paper

involves a subtlety: we are computing the average of time-dependent quantities (via the first argument in $f(n, \cdot), g(r, \cdot)$) with respect to an invariant probability on the second argument, ω . The mathematical meaning is the following:

$$E^{(sp)}[f(n, \cdot) g(r, \cdot)] = \int_{\Omega} f(n, \omega) g(r, \omega) \mu^{(sp)}(d\omega).$$

We numerically compute such quantities by generating M spike trains, denoted $\omega^{(m)}$, $m = 1 \dots M$, of length T , with the spontaneous dynamics, thus distributed according to $\mu^{(sp)}$. Then:

$$\int_{\Omega} f(n, \omega) g(r, \omega) d\mu^{(sp)}[\omega] \sim \frac{1}{M} \sum_{m=1}^M f(n, \omega^{(m)}) g(r, \omega^{(m)}). \quad (30)$$

The equality holds in the limit $M \rightarrow \infty$. Here, we chose $M = 10,000$ or $M = 100,000$. Fluctuations about the mean are ruled by the central limit theorem, so they decrease like $\frac{1}{\sqrt{M}}$ with a proportionality factor depending on the observables f and g .

5.2. Results

We consider a one-dimensional lattice of $N = 30$ neurons, separated by a lattice spacing d_x , with null boundary conditions. The connectivity is depicted in Figure 2. Each neuron excites its neighbours with a weight $W_{k\pm 1,k} = w^+ > 0$ and inhibits its second neighbours with a weight $W_{k\pm 2,k} = w^- < 0$. We compare the spontaneous activity (with a noise term and a constant term I_0 to fix the baseline activity, as in (18)) to the dynamics in the presence of Gaussian pulse, with width Δ , propagating at speed v from left to right:

$$S(x, t) = \frac{A}{\sqrt{2\pi} \Delta} e^{-\frac{1}{2} \frac{(x-vt)^2}{\Delta^2}}, \quad (31)$$

where x, t are continuous space-time variables. The neuron k is located at $x_k = k \delta$ and time is updated at each $t = n b$ where b is a time bin. In the simulations $d_x = 1 \text{ mm}$, $v = 2 \text{ mm/s}$, $b = 10 \text{ ms}$, $\Delta = 1 \text{ mm}$. The term A represent the amplitude and is variable to show the validity of the linear response theory as the amplitude of the stimulus increases. We fix $\theta = 1$, $\gamma = 0.6$, corresponding to a characteristic decay time (19) $\tau_\gamma = -\left[\frac{1}{\log \gamma}\right] \sim 2$. The memory D appearing in the summations (28), (29) was taken to be $D = 10$ from the study of correlation decay rate. This is a good compromise between the convergence of these sums and the computational time.

The constant stimulus I_0 is fixed to have a good baseline activity. We consider the activity without and with stimulus with moderate excitatory connectivity and strong inhibitory connectivity ($w^+ = 0.2$, $w^- = 2$). In Figure 3 we show the spike network activity in spontaneous activity (left) and in the presence of a moving stimulus (right). The strong inhibition is particularly visible in the presence of the stimulus.

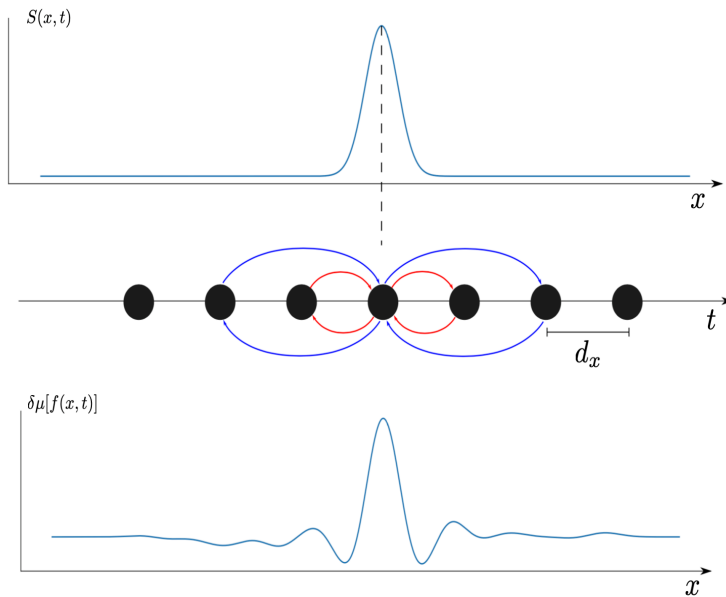


Figure 2. Connectivity pattern of the model. Neurons (black filled circles) are located into a one-dimensional lattice with spacing d_x . They are connected to nearest neighbors through excitatory connections (red) with weight w^+ and to second nearest neighbors through inhibitory connections (blue) with weight w^- . These neurons are submitted to a space-time dependent stimulus $S(x, t)$ (top line, blue trace) traveling at speed v from left to right. This modifies the average activity $\mu[f(x, t)]$ by a variation $\delta\mu[f(x, t)]$ (bottom trace).

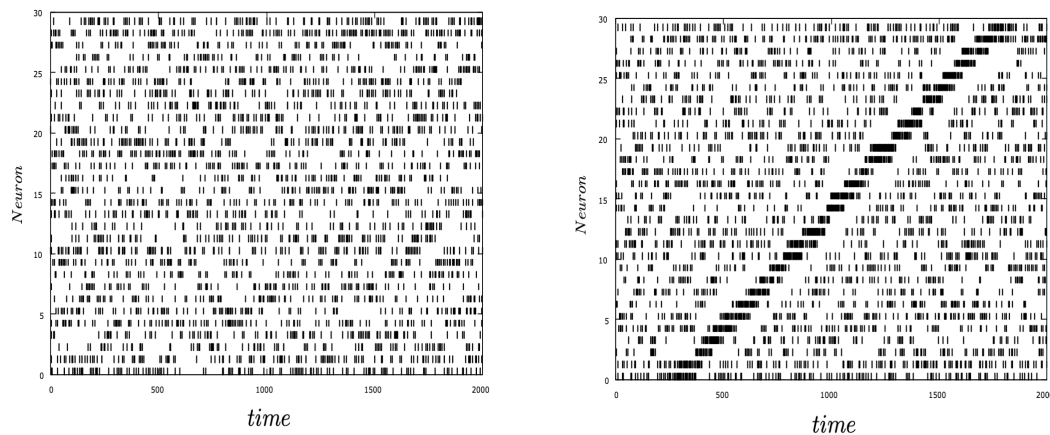


Figure 3. Spiking activity of the network. Left panel shows the spontaneous activity and right panel in the presence of a moving stimulus (31).

5.3. Linear Response for Firing Rates

We first present the results of linear response for the observable $f(n, \omega) = \omega_{k_c}(n)$, where $k_c = \frac{N}{2}$ is the index of the neuron located at the center of the lattice. Thus, $\mu[f(n, \omega)] \equiv r(k_c, t)$, the firing rate of this neuron as a function of time. In spontaneous activity it is a constant; under stimulation it depends on time. In Figure 4 we show the effect of the stimulus on the average value of this observable for different amplitudes values. One observes the combined effect of the stimulus and of the connectivity.

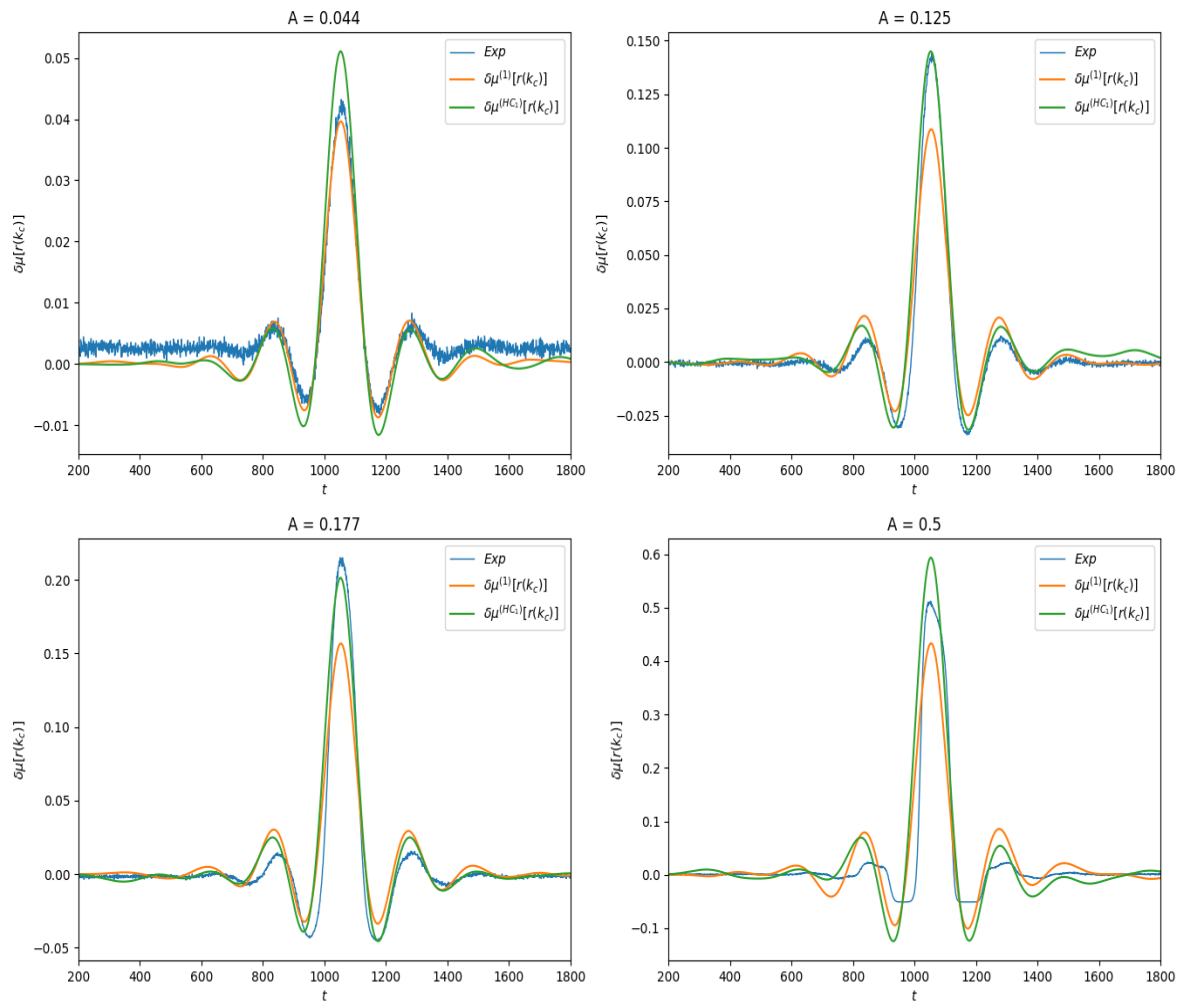


Figure 4. Linear response of $f(\omega, n) = \omega_{k_c}(n)$ for different values of stimulus amplitude A . Blue curve: Empirical average trace computed from (30). Orange: Linear response computed from Equation (28). Green: Linear response computed from Equation (29).

We next studied how correlation functions in spontaneous activity depend on space and time. One observes that they decay relatively fast with the time delay of m (Figure 5). In addition, they are multiplied by γ^m in (28), (29). Therefore the contribution to the linear response series decay exponentially fast and the series can be truncated to low order. Here we took a maximal order $D = 10$.

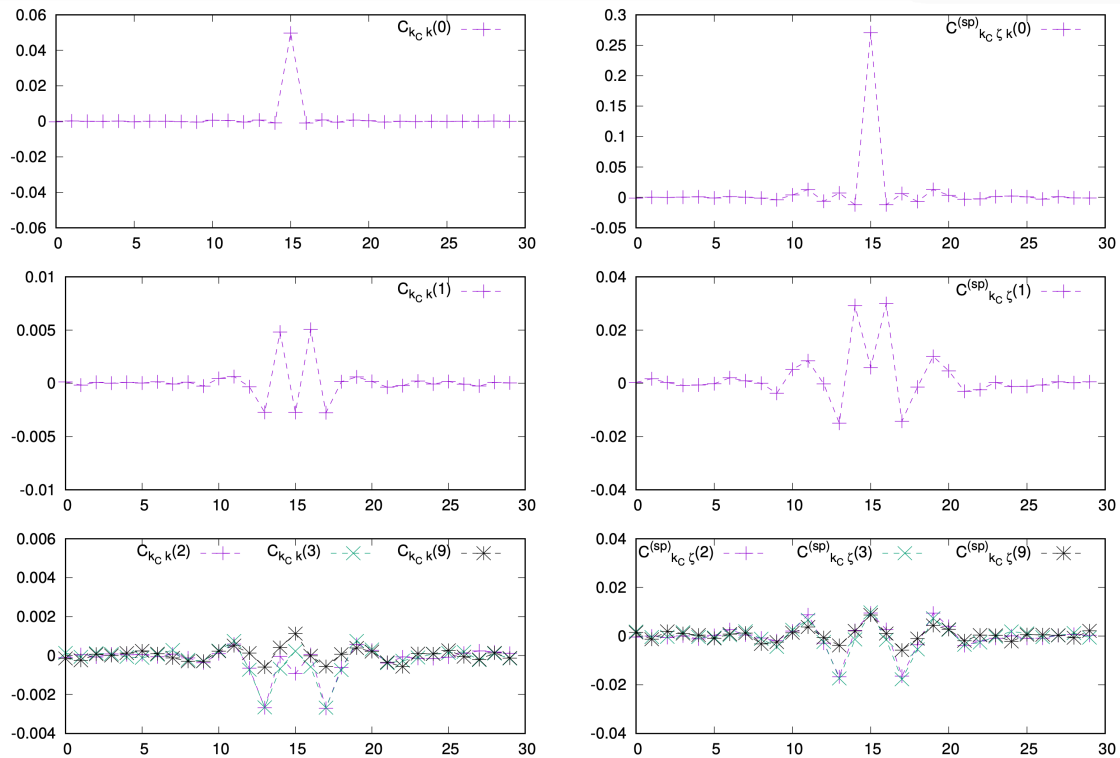


Figure 5. Correlation functions corresponding to the firing rate of the neuron $k_c = \frac{N}{2}$ as a function of the neuron index k (abscissa), for different values of the time delay m . **(Left)** correlations with stimulus. **(Right)** correlations in the spontaneous regime. Top. $m = 0$, middle $m = 1$, bottom $m = 2, 3, 9$.

Finally, as shown in Figure 4, we compute the linear response $\delta\mu^{(1)}[f(t)]$, $\delta\mu^{(HC_1)}[f(t)]$ and compare them to the response obtained by empirical averages.

5.4. Linear Response for Higher Order Observables

Here, we consider the pairwise observable $f(\omega, n) = \omega_{k_c-2}(n-3)\omega_{k_c}(n)$ where $k_c = \frac{N}{2}$. This is an example of an observable with a time delay. Neurons $k_c - 2$ and k_c mutually inhibit each other so we expect that the state of neuron $k_c - 2$ before n impact the state of neuron k_c at time n . However, the correlation between those states depends as well on the state of the other neurons.

Similarly to the previous section we have plotted in Figure 6 the empirical estimation of the linear response under the two approximations (28), (29).

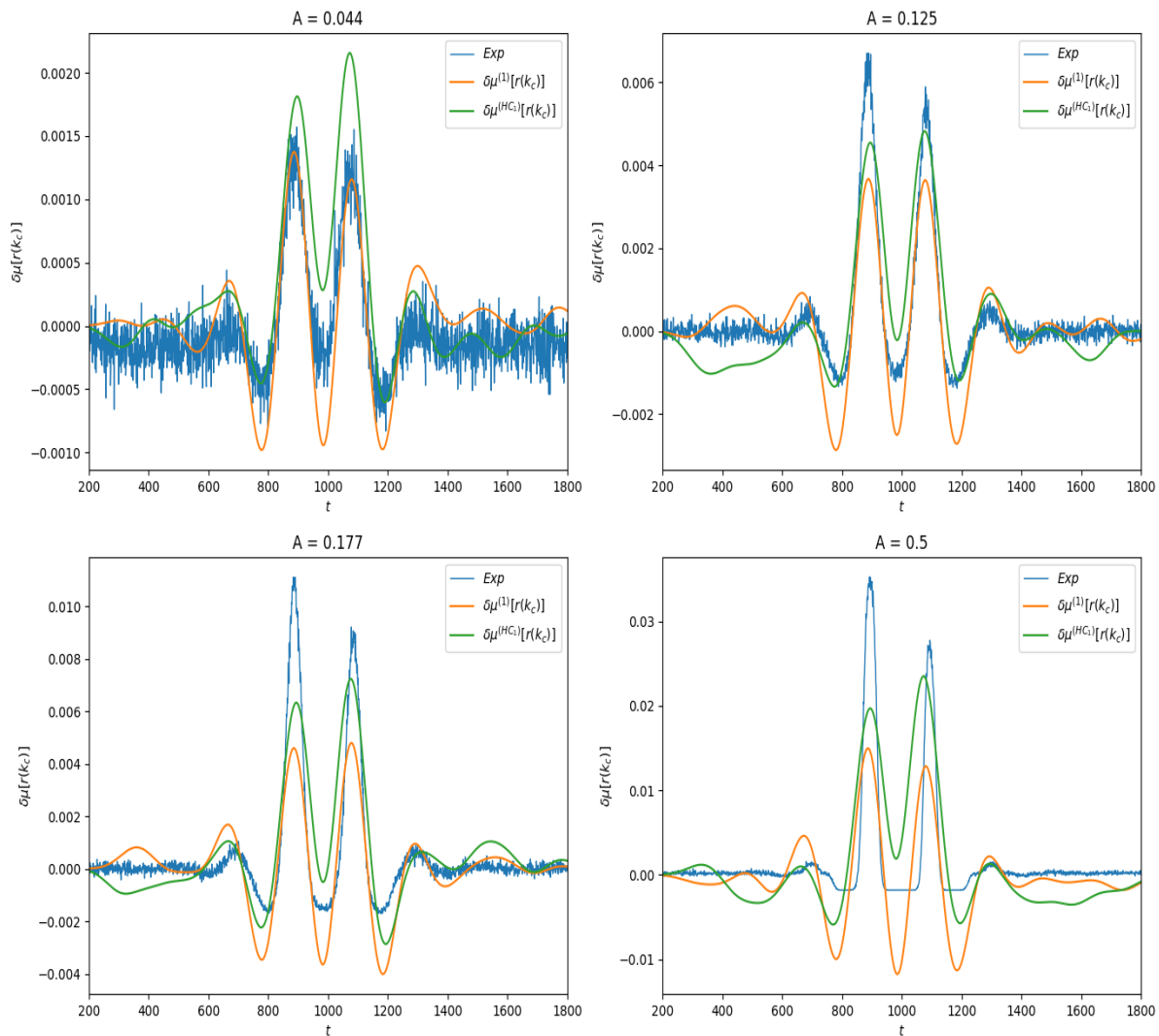


Figure 6. Linear response of the observable $f(n, \omega) = \omega_{k_c-2}(n-3)\omega_{k_c}(n)$. Here, we consider the same amplitudes and as Figure 4.

5.5. Validity of the Linear Response

The linear response is expected to hold when the stimulus amplitude is weak. What does that mean? A mathematical answer is given by Equation (27) although remaining at a rather abstract level. Here, we compute the distance:

$$d^2(\delta\mu^{(th)}[f(n)], \delta\mu^{(exp)}[f(n)]) = \sum_{n=1}^T \left(\delta\mu^{(th)}[f(n)] - \delta\mu^{(exp)}[f(n)] \right)^2, \quad (32)$$

between the theoretical curves $\delta\mu^{(th)}[f(n)]$ and the experimental curve $\delta\mu^{(exp)}[f(n)]$, as a function of the stimulus amplitude A , in the two examples of observable investigated here. Note that distance here is not normalized: it does not take into account the amplitude of the response. This explains why the distance is larger in the case of firing rates than in the delayed pairwise case, as in the latter the norm of the curve is quite smaller. The result is presented in Figure 7. As expected, the error in both cases increases with the amplitude of the stimulus, and it increases slower for (28) than for the lowest order Hammersley–Clifford expansion (29). It is interesting to see how the curves differ when A increases (see Figures 4 and 6). The empirical average curves clearly exhibit a non-linear saturation (e.g.,

firing rate cannot exceed 1) that is not reproduced by the linear response theory. This is further commented on in the discussion section.

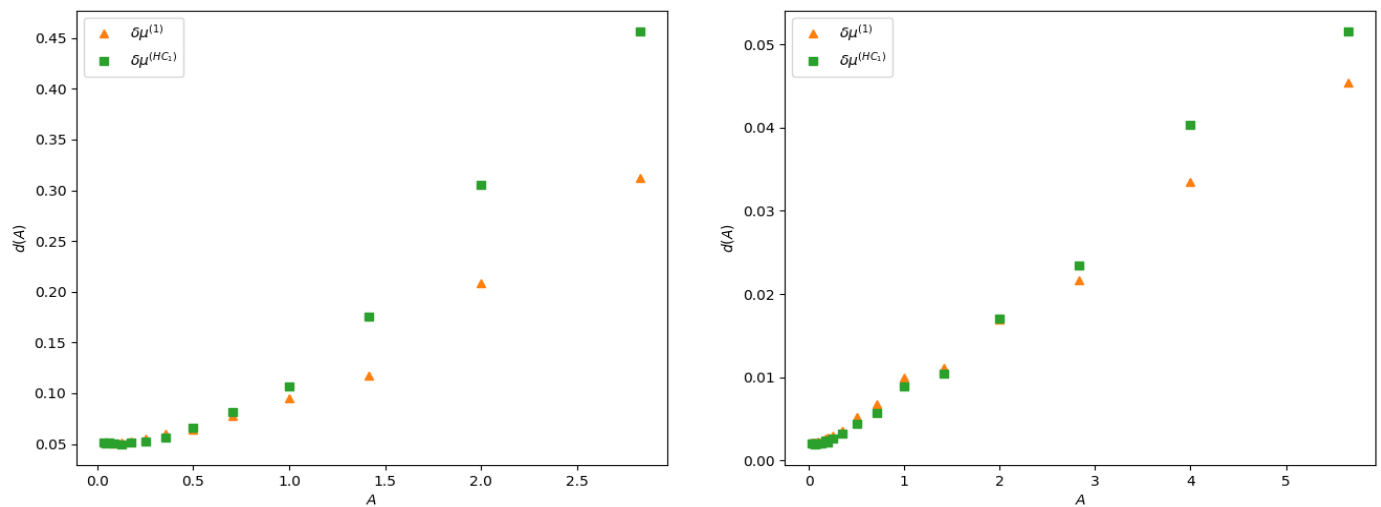


Figure 7. d^2 distance (32) between the curves $\delta\mu^{(1)}[f(n)]$, $\delta\mu^{(HC_1)}[f(n)]$ and the empirical curve, as a function of the stimulus amplitude A . The left panel shows the distance between rate curves (Section 5.3) and right panel distance between pairwise with delay (Section 5.4).

5.6. Comments on Numerical Results

With these simulations, we have been able to illustrate our results. We are able to predict the time variation of an observable under a non-stationary stimulation, from the mere knowledge of the stimulus and the spontaneous statistics. In particular, our results hold for observables with time delays, considerably enlarging the scope of linear response theories in neuronal networks. The comparison with classical fluctuation theory also emphasizes the role played by higher-order terms. Linear response requires that the stimulus has a weak enough amplitude. We see very well this effect numerically. As the amplitude A of the stimulus increases, we check the increasing discrepancy between the empirical averages and the prediction.

6. Discussion

In this paper, we have addressed the following question: How is the average of an observable $f(t, \omega)$ affected by a weak time-dependent stimulus. We studied this question in a theoretical setting, using the linear response theory and probability distributions with unbounded memory generalizing the usual definition (3) of Gibbs distributions in statistical physics courses. Our goal was to show a general mathematical formalism allowing one to handle spike correlations as a result of neuronal network activity in response to a stimulus. The most salient result of this work is that the difference of an observable average in response to a time-dependent external stimulus of weak amplitude can be computed from the knowledge of the spontaneous correlations, i.e., from the dynamics without the stimulus. This result is not surprising from a non-equilibrium statistical physics perspective (Kubo relations, fluctuation–dissipation relation [31,33]). However, to the best of our knowledge, this is the first time it has been established for spiking neuronal networks. The novelty of our approach is that it provides a consistent treatment of the expected perturbation of higher-order correlations, going in this way, beyond the known linear perturbation of firing rates and instantaneous pairwise correlations; in particular, it extends to time-dependent correlations.

In addition, we made explicit the linear response kernel in terms of the parameters determining individual networks dynamics and neuron connectivity. We have provided an explicit example of this for a well-known class of models, the LIF model. This makes explicit the role of the neuronal network structure (especially synaptic weights) in the

spiking response. As we show, the stimulus-response and dynamics are entangled in a complex manner. For example, the response of a neuron k to a stimulus applied on neuron i does not only depends on the synaptic weight W_{ki} but, in general, on all synaptic weights, because the dynamics create complex causality loops which build up the response of neuron k [8,43,44]. Our linear response formula is written in terms of the parameters of a spiking neuronal network model and the spike history of the network. Although a linear treatment may seem a strong simplification, our results suggest that already, in this case, the connectivity architecture should not be neglected. In the presence of stimuli, the whole architecture of synaptic connectivity, history and the dynamical properties of the networks are playing a role in the correlations through the perturbed potential. This agrees well with results from a recent study exhibiting an exact analytical mapping between neuronal network models and maximum-entropy models, showing that, in order to accurately describe the statistical behavior of any observable in the Maximum Entropy model, all the synaptic weights are needed, even to predict firing rates of single neurons [8].

Higher order terms can not only play a significant role in spatial terms, as shown in [63,64], but also in temporal terms. Indeed, as argued throughout this paper, neuronal interactions involve delays that have to be integrated into a model attempting to explain spike statistics [21]. Note, however, that contrary to what is usually believed, detailed balance is absolutely unnecessary in order to properly handle time correlations [65–67]. Note also that binning—which can be convenient to remove short-range time-correlations from the analysis—dramatically changes the nature of the process under investigation, rendering it non-Markovian [68].

We have also introduced the monomial expansion by providing a canonical way of decomposing the potential, describing the stationary dynamics. Moreover, the Hammersley–Clifford decomposition allows us to obtain the coefficients weighting the monomials in terms of the parameters constraining dynamics. In the case of the discrete-time LIF model, this allowed us to show the explicit dependence of the coefficients in terms of synaptic weights. Although the basis of monomials is quite huge, standard results in ergodic theory and transfer matrices/operators state that we can neglect high order terms because of the exponential correlation decay.

In the example we present we study the following question: What is the role of this lateral connectivity in motion processing in sensory neurons? Clearly, one may expect it to induce spatial and temporal correlations in spiking activity, as an echo, a trace, of the object's trajectory. These correlations cannot be read in the variations of firing rate; they also cannot be read in synchronous pairwise correlations as the propagation of information due to lateral connectivity necessarily involves delays. This example raises the question about what information can be extracted from spatio-temporal correlations in a network of connected neurons submitted to a transient stimulus.

Our results are written in terms of Kernels which can be found for any observable, generalizing the concept of receptive fields to general spatio-temporal observables beyond firing rates. This is the analogous of having “population receptive fields” which are an extension of the concept of receptive-field usually associated individual sensory neurons.

Author Contributions: conceptualization, B.C. and R.C.; methodology, B.C. and R.C.; software, B.C. and I.A.; writing—original draft preparation, B.C. and R.C.; writing—review and editing, B.C. and R.C.; visualization, B.C., R.C. and I.A.; supervision, B.C. and R.C.; project administration, B.C. and R.C.; funding acquisition, B.C. and R.C. All authors have read and agreed to the published version of the manuscript.

Funding: All authors have been supported by the MAGMA-EQA-041903 INRIA associated team. RC was supported by Fondecyt Iniciación 2018 Proyecto 11181072. BC was supported by the French government through the UCA-Jedi project managed by the National Research Agency (ANR-15-IDEX-01) and, in particular, by the interdisciplinary Institute for Modeling in Neuroscience and Cognition (NeuroMod) of the Université Côte d’Azur.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The codes and instructions required to simulate spiking data and to reproduce our numerical results can be found at <https://github.com/brincolab/Linear-response>.

Acknowledgments: We thank Arnaud Le Ny and Roberto Fernandez for helpful comments on theoretical aspects of generalized Gibbs measures and chains with complete connections.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Example Hammersley–Clifford Decomposition

The decomposition (6) is straightforward. However, we would like to give a simple, yet illustrative example. Let us consider 1 neuron at times 0 and 1. The spiking patterns are $\omega = (\omega_1(0) \omega_1(1)) = (0, 0), (1, 0), (0, 1), (1, 1)$, and, a function of these patterns takes four values: $f(0, 0) = F_0, f(1, 0) = F_1, f(0, 1) = F_2, f(1, 1) = F_3$. We have:

$$f(\omega) = F_0 (1 - \omega_1(0)) (1 - \omega_1(1)) + F_1 \omega_1(0) (1 - \omega_1(1)) \\ + F_2 (1 - \omega_1(0)) \omega_1(1) + F_3 \omega_1(0) \omega_1(1),$$

that we may rewrite in the form (6) :

$$f(\omega) = f_0 + f_1 \omega_1(0) + f_2 \omega_1(1) + f_3 \omega_1(0) \omega_1(1); \\ f_0 = F_0; f_1 = F_1 - F_0; f_2 = F_2 - F_0; f_3 = F_0 - F_1 - F_2 + F_3$$

More generally the transformation from the “function” representation (the vector F of F_i s in the example) to the monomial representation (the vector f , of f_i s) is given by the linear transformation:

$$f = Q \cdot F, \quad (A1)$$

where Q is a triangular matrix given by [45]:

$$Q_{ll'} = \begin{cases} (-1)^{d(l)-d(l')}; & l' \subseteq l \\ 0; & \text{otherwise} \end{cases}$$

Here the notation $l' \subseteq l$ means the following: To each spike block ω_0^D , one can associate a unique integer:

$$l = \sum_{k=1}^N \sum_{n=0}^D 2^{nN+k-1} \omega_k(n), \quad (A2)$$

called the index of the block. We define the block inclusion \sqsubseteq on $\Omega_{N,R}$: $\omega_0^D \sqsubseteq \omega'^0_{-D}$ if $\omega_k(n) = 1 \Rightarrow \omega'_k(n) = 1$ (all bits ‘1’ in ω_0^D are bits ‘1’ in ω'^0_{-D}), with the convention that the block of degree 0 is included in all blocks. By extension $l' \subseteq l$ means that all bits ‘1’ in the block corresponding to the integer l are included in the block corresponding to l' [8].

The result (A1) shows that the coefficient of a monomial is the linear combination of function values where the number of terms in the combination increases with the monomial degree.

Appendix B. Monomials Decomposition

Appendix B.1. Linear Response in the Finite-Range Potential Approximation

Although $\phi, \phi^{(sp)}$ have infinite range, their memory dependence can decay fast, typically exponentially. This property is independent of the application of a stimulus: it holds for ϕ and $\phi^{(sp)}$ as well, hence for $\delta\phi$. In this case, the infinite range potentials can be approximated by a finite-range one. We use here a classical result in ergodic theory: a potential ϕ with an exponentially decaying variation (11) (more generally, so-called regular potential) can also be approximated by a finite-range potential $\phi^{(R)}$ in the sup norm where $\|\phi - \phi^{(R)}\| \leq C\Theta^R$, for some $0 < \Theta < 1$ [69]. Equivalently the chain with infinite memory

can be replaced by a Markov chain of memory depth D , $D = R + 1$. Therefore, $\delta\phi$ can be approximated by a range- R potential $\delta\phi^{(R)}$. Using the monomial decomposition (6), we have:

$$\delta\phi(r, \omega) \sim \sum_{l'} \delta\phi_l(r) m_l(\omega_{r-D}^r).$$

In this setting, (16) becomes:

$$\delta^{(1)}[f(t)] = \sum_{r=n_0+1}^{n=[t]} \sum_{l,l'} f_l(t) \delta\phi_{l'}(r) \mathcal{C}^{(sp)}[m_l(\omega_{n-D}^n), m_{l'}(\omega_{r-D}^r)].$$

However, $\mu^{(sp)}$ is stationary by assumption. Hence the correlation $\mathcal{C}^{(sp)}[m_l(\omega_{n-D}^n), m_{l'}(\omega_{r-D}^r)]$ only depends on the two monomials $m_l, m_{l'}$ and on the time lag $n - r$, i.e., $\mathcal{C}^{(sp)}[m_l(\omega_{n-D}^n), m_{l'}(\omega_{r-D}^r)] = \mathcal{C}^{(sp)}[m_l \circ \sigma^{n-r}, m_{l'}]$, that we write $C_{l,l'}(n - r)$ to alleviate notations. Thus:

$$\delta^{(1)}[f(t)] = \sum_{r=-\infty}^{n=[t]} \sum_{l,l'} f_l(t) C_{l,l'}(n - r) \delta\phi_{l'}(r),$$

so that the linear response is a linear decomposition of monomial correlations computed at equilibrium. We can write this in a more compact form introducing the L -dimension vectors $f(t) = (f_l(t))_{l=1}^L$, $\delta\phi(r) = (\delta\phi_l(r))_{l=1}^L$ and the matrix $\mathcal{C}(n - r) = (C_{l,l'}(n - r))_{l,l'=1}^L$. We write $\delta^{(1)}[f(t)]$ in the form:

$$\delta^{(1)}[f(t)] = \sum_{r=-\infty}^{n=[t]} \langle f(t) | \mathcal{C}(n - r) \cdot \delta\phi(r) \rangle, \quad (\text{A3})$$

where $\langle | \rangle$ denotes the standard scalar product.

This has three important consequences.

Appendix B.2. The Convolution Kernel

For a time independent observable $f(\omega)$ we introduce $K_f(m)$ the $L \times L$ matrix with entries:

$$K_{f,l,l'}(m) = f_l C_{l,l'}(m),$$

so that:

$$\delta^{(1)}[f(t)] = \sum_{r=-\infty}^{n=[t]} \sum_{l,l'} K_{f,l,l'}(n - r) \delta\phi_{l'}(r),$$

which is a discrete convolution. It is close to the form (1) with the difference that the time is discrete, coming from our spike trains discretization. The stimulus, explicit in (1) is here hidden in $\delta\phi(r)$. We give an illustration of this in the next section. However, the fundamental result is that the convolution kernel defined this way is a linear combination of monomial correlations functions. This has, therefore, the form of a Kubo equation where the monomials play the role of the physical quantities introduced in Section 2.1. As mentioned above, the main and essential difference is that, in contrast to Physics, we have no a priori idea which monomials are the most important (except straightforward arguments on the decaying probability of high order monomials). There is no known principle to guide the choice.

References

- Abbott, L.; Dayan, P. The effect of correlated variability on the accuracy of a population code. *Neural Comput.* **1999**, *11*, 91–101.
- Rieke, F.; Warland, D.; van Steveninck, R.d.; Bialek, W. *Spikes, Exploring the Neural Code*; The M.I.T. Press: Cambridge, MA, USA, 1996.
- Masquelier, T.; Guyonneau, R.; Thorpe, S. Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. *PLoS ONE* **2008**, *3*, e1377.
- Schneidman, E.; Segev, M.J.B.I.I.R.; Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **2006**, *440*, 1007–1012.
- Tkačik, G.; Marre, O.; Mora, T.; Amodei, D.; Berry, M.J., II; Bialek, W. The simplest Maximum Entropy model for collective behavior in a neural network. *J. Stat. Mech.* **2013**, *2013*, P03011.
- Pillow, J.; Paninski, L.; Uzzell, V.; Simoncelli, E.; Chichilnisky, E.J. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *J. Neurosci.* **2005**, *25*, 11003–11013.
- Vasquez, J.C.; Palacios, A.; Marre, O.; Berry, M.J., II; Cessac, B. Gibbs distribution analysis of temporal correlation structure on multicell spike trains from retina ganglion cells. *J. Physiol. Paris* **2012**, *106*, 120–127.
- Cofré, R.; Cessac, B. Exact computation of the maximum-entropy potential of spiking neural-network models. *Phys. Rev. E* **2014**, *89*, 052117.
- deCharms, R.; Merzenich, M. Primary cortical representation of sounds by the coordination of action potentials. *Nature* **1996**, *381*, 610–613.
- Alonso, J.; Usrey, W.; Reid, W. Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature* **1996**, *383*, 815–819.
- Marre, O.; Botella-Soler, V.; Simmons, K.D.; Mora, T.; Tkačik, G.; Berry, M.J., 2nd. High accuracy decoding of dynamical motion from a large retinal population. *PLoS Comput. Biol.* **2015**, *11*, e1004304.
- Tsodyks, M.; Kenet, T.; Grinvald, A.; Arieli, A. Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* **1999**, *286*, 1943–1946.
- Dayan, P.; Abbott, L.F. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*; MIT Press: Cambridge, MA, USA, 2001.
- Gerstner, W.; Kistler, W. *Spiking Neuron Models*; Cambridge University Press: Cambridge, UK, 2002.
- Nirenberg, S.; Latham, P. Decoding neuronal spike trains: how important are correlations. *Proc. Nat. Acad. Sci. USA* **2003**, *100*, 7348–7353.
- Sompolinsky, H.; Yoon, H.; Kang, K.; Shamir, M. Population coding in neuronal systems with correlated noise. *Phys. Rev. E* **2001**, *64*, 051904.
- Brunel, N.; Hakim, V. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.* **1999**, *11*, 1621–1671.
- Trousdale, J.; Hu, Y.; Shea-Brown, E.; Josic, K. Impact of network structure and cellular response on spike time correlations. *PLoS Comput Biol.* **2012**, *8*, e1002408.
- Lindner, B.; Doiron, B.; Longtin, A. Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. *Phys. Rev. E* **2005**, *72*, 061919.
- Ostojic, S.; Brunel, N.; Hakim, V. How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *J. Neurosci.* **2009**, *33*, 10234–10253.
- Marre, O.; Amodei, D.; Deshmukh, N.; Sadeghi, K.; Soo, F.; Holy, T.E.; Berry, M.J., 2nd. Mapping a complete neural population in the retina. *J. Neurosci.* **2012**, *32*, 14859–14873.
- Ferrea, E.; Maccione, A.; Medrihan, L.; Nieuws, T.; Ghezzi, D.; Baldelli, P.; Benfenati, F.; Berdondini, L. Large-scale, high-resolution electrophysiological imaging of field potentials in brain slices with microelectronic multielectrode arrays. *Front. Neural Circuits* **2012**, *6*, 80.
- Croner, L.; Purpura, K.; Kaplan, E. Response variability in retinal ganglion cells of primates. *PNAS* **1993**, *90*, 8128–8130.
- Shadlen, M.; Newsome, W. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* **1998**, *18*, 3870–3896.
- Ferrari, U. Learning Maximal Entropy Models from finite size datasets: a fast Data-Driven algorithm allows to sample from the posterior distribution. *Phys. Rev. E* **2016**, *94*, 023301.
- Botella-Soler, V.; Deny, S.; Martius, G.; Marre, O.; Tkačik, G. Nonlinear decoding of a complex movie from the mammalian retina. *PLoS Comput. Biol.* **2018**, *14*, e1006057.
- Nghiem, T.; Telenczuk, B.; Marre, O.; Destexhe, A.; Ferrari, U. Maximum-entropy models reveal the excitatory and inhibitory correlation structures in cortical neuronal activity. *Phys. Rev. E* **2018**, *98*, 012402.
- Ermentrout, B.; Terman, D. *Mathematical Foundations of Neuroscience*, 1st ed.; Springer: Berlin, Germany, 2010.
- Izhikevich, E.M. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*; The MIT Press: Cambridge, MA, USA, 2007.
- Onicescu, O.; Mihoc, G. Sur les chaînes statistiques. *C. R. Acad. Sci. Paris* **1935**, *200*, 511–512.
- Kubo, R. The fluctuation-dissipation theorem. *Rep. Prog. Phys.* **1966**, *29*, 255–284.
- Kubo, R. Statistical-mechanical theory of irreversible processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems. *J. Phys. Soc.* **1957**, *12*, 570–586.

33. Ruelle, D. Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics. *J. Statist. Phys.* **1999**, *95*, 393–468.
34. Cessac, B. A discrete time neural network model with spiking neurons ii. dynamics with noise. *J. Math. Biol.* **2011**, *62*, 863–900.
35. Hammersley, J.; Clifford, P. Markov fields on finite graphs and lattices. **1971**. Unpublished.
36. Paninski, L. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Comput. Neural Syst.* **2004**, *15*, 243–262.
37. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620.
38. Ruelle, D. A variational formulation of equilibrium statistical mechanics and the Gibbs phase rule. *Comm. Math. Phys.* **1967**, *5*, 324–329.
39. Georgii, H.-O. *Gibbs Measures and Phase Transitions*. De Gruyter: Berlin, Germany, 2011; Volume 9.
40. Bellac, M.L.; Mortessagne, F.; Batrouni, G.G. *Equilibrium and Non-Equilibrium Statistical Thermodynamics*; Cambridge University Press: Cambridge, UK, 2004.
41. Amari, S. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* **1977**, *27*, 77–87.
42. Wilson, H.; Cowan, J. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* **1973**, *13*, 55–80.
43. Cessac, B.; Sepulchre, J.A. Stable resonances and signal propagation in a chaotic network of coupled units. *Phys. Rev. E* **2004**, *70*, 056111.
44. Cessac, B.; Sepulchre, J.A. Transmitting a signal by amplitude modulation in a chaotic network. *Chaos* **2006**, *16*, 013104.
45. Moussouris, J. Gibbs and Markov random systems with constraints. *J. Stat. Phys.* **1974**, *10*, 11–33.
46. Seneta, E. *Non-Negative Matrices and Markov Chains*; Springer: Berlin, Germany, 2006.
47. Gantmacher, F.R. *The Theory of Matrices*; AMS Chelsea Publishing: New York, NY, USA, 1964.
48. Walters, P. A variational principle for the pressure of continuous transformations. *Amer. J. Math.* **1975**, *97*, 937–971.
49. Bowen, R. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, 2nd ed.; Lecture Notes in Mathematics 470; Springer: Berlin, Germany, 1975.
50. Galves, A.; Locherbach, E. Infinite Systems of Interacting Chains with Memory of Variable Length-A Stochastic Model for Biological Neural Nets. *J. Stat. Phys.* **2013**, *151*, 896–921.
51. Reynaud-Bouret, P.; Rivoirard, V.; Grammont, F.; Tuleau-Malot, C. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *J. Math. Neurosci.* **2014**, *4*, 1–41.
52. Vidybida, A.K. Activity of excitatory neuron with delayed feedback stimulated with Poisson stream is non-Markov. *J. Stat. Phys.* **2015**, *160*, 1507–1518.
53. Fernandez, R.; Maillard, G. Chains with complete connections : General theory, uniqueness, loss of memory and mixing properties. *J. Stat. Phys.* **2005**, *118*, 555–588.
54. Kozlov, O. Dynamics of encoding in a population of neurons. *J. Gen. Physiol.* **1972**, *59*, 734–766.
55. Ny, A.L. Introduction to (generalized) Gibbs Measures. *Ensaio Mat.* **2008**, *15*, 1–126.
56. Fernandez, R.; Gallo, S. Regular g-measures are not always Gibbsian. *Electron. Commun. Probab.* **2011**, *16*, 732–740.
57. Cofré, R.; Cessac, B. Dynamics and spike trains statistics in conductance-based integrate-and-fire neural networks with chemical and electric synapses. *Chaos Solitons Fractals* **2013**, *50*, 13–31.
58. Cessac, B.; Viéville, a.T. On Dynamics of Integrate-and-Fire Neural Networks with Adaptive Conductances *Front. Comput. Neurosci.* **2008**, *2*, 2.
59. Cessac, B. Statistics of spike trains in conductance-based neural networks: Rigorous results. *J. Math. Neurosci.* **2011**, *1*, 1–42.
60. Schöpf, H.M.; Supancic, P.H. On Bürmann’s Theorem and Its Application to Problems of Linear and Nonlinear Heat Transfer and Diffusion. *Math. J.* **2014**, *16*, doi:10.3888/TMJ.16-11.
61. Stimberg, M.; Brette, R.; Goodman, D.F.M. Brian 2, an intuitive and efficient neural simulator *eLife* **2019**, *8*, e47314.
62. Cessac, B.; Kornprobst, P.; Kraria, S.; Nasser, H.; Pamplona, D.; Portelli, G.; Viéville, T. PRANAS: A New Platform for Retinal Analysis and Simulation *Front. Neuroinform.* **2017**, *11*, 49.
63. Ganmor, E.; Segev, R.; Schneidman, E. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *PNAS* **2011**, *108*, 9679–9684.
64. Ganmor, E.; Segev, R.; Schneidman, E. The Architecture of Functional Interaction Networks in the Retina. *J. Neurosci.* **2011**, *31*, 3044–3054.
65. Cofré, R.; Maldonado, C. Information Entropy Production of Maximum Entropy Markov Chains from Spike Trains. *Entropy* **2018**, *20*, 34.
66. Cofré, R.; Maldonado, C.; Rosas, F. Large Deviations Properties of Maximum Entropy Markov Chains from Spike Trains. *Entropy* **2018**, *20*, 573.
67. Cofré, R.; Rosas, L.V.A.F. An Introduction to the Non-Equilibrium Steady States of Maximum Entropy Spike Trains. *Entropy* **2019**, *21*, 884.
68. Cessac, B.; Ny, A.L.; Locherbach, E. On the mathematical consequences of binning spike trains. *Neural Comput.* **2017**, *29*, 146–170.
69. Ledrappier, F. Principe variationnel et systemes dynamiques symboliques. *Z. Wahrscheinlichkeitstheorie Verw. Geb.* **1974**, *30*, 185–202.